Journal of
Clinical Psychology

JANUARY 1958
Volume XIV Number 1

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Editorial Office: 5 Pearl Street
Brandon, Vermont

JERRY W. CARTER, JR.
U. S. Public Health Service

HOWARD F. HUNT
University of Chicago

WILLIAM A. HUNT
Northwestern University

ELAINE F. KINDER
New York State Department of Mental Hygiene

MAURICE LORR
Veterans Administration Washington, D. C.

SAUL B. SELLS
School of Aviation Medicine Randolph Field, Texas

ROBERT I. WATSON
Northwestern University

ELEANOR C. THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street · · Brandon, Vermont

COMPANION VOLUMES

PRINCIPLES OF PSYCHOLOGICAL EXAMINING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $7.50

PRINCIPLES OF PERSONALITY COUNSELING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $6.00

SYSTEMATIC: The first comprehensive system of diagnosis and psychotherapy based solidly on basic psychological science.

THEORETICAL: Presenting a comprehensive theory of the nature of personality integration with special emphasis on the differential diagnosis and modification of the primary etiologic factors which actually organize socially significant behaviors.

ECLECTIC: All pertinent theories and facts in basic science psychology and clinical psychological science are related in a systematic applied integrative psychology.

COMPLETE: For all teaching and reference purposes.

ADVANCED: Both of these books present a new orientation to diagnosis, counseling and psychotherapy which is more than just a rehash of the literature. Principles of Psychological Examining is more than a textbook of psychometrics since tests and measures are presented only as they relate to more fundamental problems of psychological diagnosis. Similarly, Principles of Personality Counseling presents thorough discussions of the indications and contraindications for applying all known methods of psychotherapy after a primary diagnosis of factors requiring modification has been accomplished.

ORDER NOW

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

JOURNAL OF CLINICAL PSYCHOLOGY

VOL. XIV JANUARY 1958 No. 1

CONTENTS

INTELLECTUAL PERFORMANCE

The Character and Degree of Psychopathology in Relation to Sustained Intellectual Performance. ROBERT ROESSLER and LEIGH ROBERTS - - 3

Spontaneous Fluctuations in the Severity of Illness of Hospitalized Lobotomized Patients. THOMAS E. HANLON, CAROLYN S. SHEETS and ALBERT A. KURLAND - 6

Differences in Some Personal Characteristics of "Rigid" and "Flexible" Individuals. K. WARNER SCHAIE - - - - - - - - - - - - - - - - 11

Further Studies on the Bender-Gestalt Test and the Digit Span Test as Measures of Recall. ALEXANDER TOLOR - - - - - - - - - - - - - 14

Age, Intelligence, Aphasia and the Spiral After-Effect in an Epileptic Population. BERNARD S. AARONSON - - - - - - - - - - - - - - - - 18

Wechsler-Bellevue Changes over Time. STEWART G. ARMITAGE and DAVID PEARL - 22

A comparison of the Bender-Gestalt Reproductions of Delinquents and Non-Delinquents. EDWIN ZOLIK - - - - - - - - - - - - - - - - 24

DIAGNOSTIC PROCESSES: GENERAL

PEAQ: A Measure of Psychopathic Behavior. MORTIMER H. APPLEZWEIG, ANDREW S. DIBNER and RAYMOND L. OSBORNE - - - - - - - - - - 26

The Hill-Monroe Inventory for Predicting Acceptability for Psychotherapy in the Institutionalized Drug Addict. JACK J. MONROE and HARRIS E. HILL - 31

A New Sorting Technique for Diagnosing Brain Damage. JOHN ALTROCCHI and B. G. ROSENBERG - 36

An Evaluation Scale for Shock Therapy. MARVIN J. FELDMAN - - - - - - 41

A Validation of Measures that Predict the Efficacy of Shock Therapy. DONALD K. PUMROY and WILLIAM S. KOGAN - - - - - - - - - - - - - 46

An Evaluation of the Grassi Test for Organic Involvement. MICHAEL HIRT 48

DIAGNOSTIC PROCESSES: PROJECTIVE TECHNIQUES

An Evaluation of the "Organic Signs" in the H-T-P Drawings. VYTAUTAS J. BIELIAUSKAS and SANDRA L. KIRKHAM - - - - - - - - - - - - - - 50

Human Figure Drawings by Younger and Older Adults. IRVING LORGE, JACOB TUCKMAN and MICHAEL B. DUNN - - - - - - - - - - - - - - 54

Manifest Content and Interpretive Meaning of Verbal Intelligence Test Responses. NORMAN TALLENT - - - - - - - - - - - - - - - - - - 57

Rorschach *M* Responses and Intelligence. ROBERT SOMMER - - - - - - 58

Global Evaluation of Rorschach Performance Versus Scores: Sex Differences in Rorschach Performance. T. W. RICHARDS and DAVID C. MURRAY 61

Utilization of O'Reilly's Objective Rorschach as a Screening Test for State Colony Job Applicants. MARY MARTHA MURPHY - - - - - - - - - - 65

Movement in the Fantasy of Brain-Injured (Cerebral Palsy) Children. T. W. RICHARDS - 67

CONTENTS—Continued

DIAGNOSTIC PROCESSES: MMPI AND RELATED SCALES

Differentiating Alcoholics from Normals and Abnormals with the MMPI. DONALD P. HOYT AND GORDON M. SEDLACEK - - - - - - - - - - 69

The Effect on Self Description of Combining Anxiety and Hostility Items on a Single Scale. MARVIN J. FELDMAN AND SAUL M. SIEGEL - - - - 74

Face Validity in the *D* Scale of the MMPI. RAYMOND J. MCCALL - - - - 77

The Taylor Manifest Anxiety Scale in Differential Diagnosis. HAROLD RUBIN and ALAN H. TOWNSEND - - - - - - - - - - - - - - - - 81

Prediction of Rehospitalization using the MMPI. PETER F. BRIGGS - - - 83

Note on a Rapid Scoring Procedure for the Card Form MMPI. GEORGE T. WILCOX - 85

STATISTICS FOR THE CLINICIAN

The Clinical Application of Sequential Analysis. ARNOLD TREHUB and ISIDOR W. SCHERER - - - - - - - - - - - - - - - - - - - 86

An Approximation Table for Determining Sample Size by Means of the Sign Test. MOSS L. RAWN - - - - - - - - - - - - - - - - - - 90

Construction of a Table of the Significance of the Difference between Verbal and Performance IQ's on the WAIS and the Wechsler-Bellevue. WOLF P. WOLFENSBERGER - 92

The Effect of a Recorded Voice on the Drawing of a Geometric Design as a Predictor of Social Behavior. HOWARD J. SHEAR - - - - - - - - - 93

A Note on the Reliability of the L-M Fergus Falls Behavior Rating Scale. LLOYD K. SINES - 95

Editorial Opinion - 99

Book Notices - 101

GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Special sections in each issue will be devoted to news, comments, and letters to the editors on matters pertinent to the profession. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

ORIGINAL ARTICLES are published only with the understanding that they are contributed exclusively to the *Journal of Clinical Psychology*. The publishers are not responsible for statements made or opinions expressed by contributors in articles printed in these columns. Copyrights cover publication in the *Journal of Clinical Psychology* and articles may not be reproduced without permission of the publishers.

THE STYLE of all manuscripts including bibliographies must be in accordance with the instructions given in: Anderson, J. E., and Valentine, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 41, 345-376. References should be arranged alphabetically according to the directions given above. Full address of the author should appear somewhere in the article, preferably at the end.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscribers' expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $8.50; Canadian subscriptions $9.00; foreign subscriptions $9.50 including postage; single copies $2.50. Entered as second class matter January 15, 1945, at the post office at Burlington, Vermont, under the Act of March 3, 1879. Reentered at the post office at Brandon, Vermont. Additional entry at the post-office at Rutland, Vermont. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont.

Copyright, 1958, Frederick C. Thorne, M. D.

THE CHARACTER AND DEGREE OF PSYCHOPATHOLOGY IN RELATION TO SUSTAINED INTELLECTUAL PERFORMANCE

ROBERT ROESSLER AND LEIGH ROBERTS

University of Wisconsin Medical School

Problem

Psychiatrists are repeatedly confronted with patient complaints of impaired intellectual functioning. This is particularly true in a college setting, where student patients almost universally complain of inability to concentrate, lowered academic performance and anxiety over examinations. It is probably safe to predict that most psychiatrists, asked to formulate hypotheses concerning the effect of psychopathology on intellectual performance, would predict impaired performance proportionate to the character and degree of pathology. The present study is designed to test these subjective impressions on a college population.

A review of the literature reveals two groups of pertinent studies. One group, dealing with the effect of variation in external stress on intellectual performance, has been reviewed comprehensively by Lazarus, et al[6]. Most of this work suffers from an almost total neglect of variables which are a function of the internal organization of the individuals tested. Consequently, individual differences tend to be one of the main findings in studies of this type.

The second group of studies, alert to the foregoing error, have attempted to correlate intellectual performance with internal variables as defined by various psychiatric groups. Hunt[5] has summarized these findings up to 1944. Greenfield[3], more recently, compared a group of college normals with a group of neurotic patients with respect to their performance on certain intellectual tasks while "frustrated" and found no significant differences between the two groups. Fey[1] found clearcut differences between schizophrenics and controls on the Wisconsin Card Sorting Test under different conditions of personal, impersonal, and physical stress. In general, in the studies of this type, schizophrenics are the only pathological group who tend to show consistent differences from controls on intellectual tasks. All of these studies were carried out under laboratory conditions and, to the clinician, there may be considerable danger in generalizing the conclusions reached in them to behavior outside the laboratory situation. Moreover, the tests of performance were, in every instance, sharply discrete temporally, *i.e.*, the performances were measured only within relatively brief spans of time. To the present authors' knowledge only one group of investigators has heretofore measured intellectual performance in a psychopathological group outside the laboratory and over a relatively extended period. Matarazzo, *et al*[7] have correlated academic performance over a year's period with various degrees of anxiety and found no significant correlation. The present study utilizes the same technique on a broader psychopathological sample.

Method

The population selected for study was all of the undergraduate students applying for psychiatric assistance in the Student Health Psychiatry Section of the Department of Psychiatry of the University of Wisconsin in the four year period of 1950-1953 inclusive. From this group all those with no psychiatric diagnosis were excluded, leaving a population of over 1000. From this population, in turn, the following five experimental groups were randomly selected: (a) 122 patients without regard to diagnosis, (b) 39 personality disorders, characterized by "acting out" behavior, (c) 44 obsessive-compulsive neurotics, (d) 43 incipient psychotics and psychotics, and (e) 59 persons diagnosed as "adolescent reaction", in whom it was felt by the examiners there was no serious pathology. On each of these samples the mean academic grade-point of the group was calculated. The mean period on which the grade-point was calculated for the experimental groups varied from 2.8 to 3.6 academic years. A comparison of the variances within each group was made. The mean

grade-point of each of these samples was then compared with the mean grade-point of 100 unselected persons from the total population of the university during the same period. In addition, where significant differences were found, the influence of intelligence in accounting for the difference was tested by comparing the experimental group with the control group on the mean score achieved on the American Council on Education Test.

Hypotheses

For statistical purposes all hypotheses were framed in terms of the null hypothesis, *i.e.*, that there would be no significant differences when each of the experimental samples was compared with the normal control sample. However, the authors' hypotheses may be framed in positive terms as follows:

(1) *Obsessive-compulsives* will perform significantly better academically than will a random sample of the total student population. This expectation was based upon the personality features of intellectualization, compulsiveness, and strong achievement needs which characterize this group of disorders.

(2) *Personality disorders* characterized by "acting out" will perform significantly poorer than the sample of the total student population. This was anticipated because it was felt that this group, under the stress of impending examinations, would discharge their anxiety tangentially and would tend to study less and value academic achievement less than would a group of normals.

(3) *Psychotics and incipient psychotics* will perform significantly poorer than "normals". This was anticipated because of the clinical impressions of interference with thought processes, overwhelming anxiety, and behavioral disorganization.

(4) *Adolescent reactions* will not differ significantly from "normals". It was felt that this group, being minimally sick from the psychiatric frame of reference, would be roughly equatable to "normals".

(5) A random sample of the total psychopathological population will not differ from "normals". This was anticipated because of the earlier hypotheses that performance in the psychopathological groups would be enhanced in some instances and impaired in others so that these opposing trends would cancel out differences in the total group.

Results

The null hypothesis cannot be rejected except for the obsessive-compulsive group. The latter group performed significantly better than the normal controls and this difference was significant at better than the 1% level of confidence. When the obsessive-compulsive group was compared with the normal on the mean performance on the American Council of Education Test, no significant difference was found; hence, greater intelligence cannot account for the enhanced performance of this group. Analysis of variance within the groups revealed no significant differences.

Discussion

Certain limitations in the design of this study should be noted. First, the pathological populations concerned are those who sought assistance; whether similar pathological populations who did *not* seek help would differ cannot be answered. Second, it might be argued that decrements in academic performance might have been revealed had the period of measurement been restricted to the period of clinical contact, during which the patients concerned were presumably most anxious. However, Matarazzo, *et al*[7] have previously demonstrated that anxiety bears no relationship to intellectual performance of this character.

With the foregoing reservations, the results of this study tend to confirm the results of laboratory experimentation, *i.e.*, that character and degree of psychopathology are not predictive of intellectual performance. The exception to this generalization in this study, that of the obsessive-compulsive group, is actually in the direction of enhanced performance. For the authors, the finding that the existence

of psychosis or incipient psychosis and "acting out" personality disorders does not impair academic performance is surprising and contrary to subjective clinical impressions. How can one view such results most usefully?

Hartman [4] and Rapaport [8], in discussing the sparing of certain ego functions in psychopathology have proposed the concept of "conflict-free ego spheres". In terms of this hypothesis it might be speculated that the intellectual performances measured in this study are at least comparatively free of conflict and spared the encroachment which is attendant to other areas of behavior in emotionally sick persons. Such an explanation, however, is not congruent with innumerable reports by patients in this study of conflict attendant to academic performance of the obsessive-compulsive group.

Greenfield [2] has made similar observations on the concept of "conflict-free" ego spheres in noting the importance of such non-intellectual factors as motivation and attention in performing intellectual tasks. He has proposed the terms "syntonic ego-involvement" and "dystonic ego-involvement" to designate respectively those behaviors which are or are not consonant with the standards of the ego. This theory is certainly attractive when applied to the obsessive-compulsive group, where its implication is that studying and examination behaviors are more syntonic to the obsessive-compulsive ego than to the "normal" ego. This interpretation is in keeping with clinical observation that for the obsessive-compulsive personality, repetitious self-assessment is an important defense. Moreover, successful examination performance probably places a premium upon personality attributes characteristic of the obsessive-compulsive— ritualistic repetition (of facts to be memorized), obsessive concern for detail, great skills in conceptualization, abstraction and verbalization. Finally, the value placed on academic achievement by the obsessive-compulsive seems clinically to be greater than the value to "normals"; the importance of such value in motivating study is probably important.

When this theory is applied to the remaining pathological groups, its usefulness is questionable. While the behaviors measured might, in part, be as useful defensively to these groups as to the obsessive-compulsives and hence ego-syntonic, the clinician would anticipate *a priori* from the reports of patients that they are largely dystonic and threatening; hence it might be expected that performance would be impaired. We can only speculate that, while the threat of these behaviors may be anxiety-provoking, the anxiety provoked is not sufficient to disorganize behavior; indeed, it may be facilitory in some degree. Were the variability of performance in these groups significantly greater than in the normal, it would tend to lend some support to this view. Unfortunately, this cannot be demonstrated in this study.

The authors believe that this work underscores the need for further objective delineation of the ways in which ego functions are affected by psychopathology. It is all too easy for the clinician to mistake phenomenologic patient reports for objectively verifiable behavior. While the usefulness of such information to the clinician may be minimal (concerned as he is largely with *experienced* distress in his patients, not so greatly with their behavior *per se*) it may be helpful to other groups necessarily concerned with psychology of man. Educators, for example, may well pause before assuming too quickly that psychopathology, like some organic pathology, constitutes an adequate explanation for lowered academic performance.

Summary

This paper reports the effects of psychopathology in a college under-graduate population on scholastic performance over a period of several years. Five experimental groups were randomly selected from among undergraduate students seen in the Student Health Psychiatry Section of the University. These were classified as personality disorders characterized by "acting out" behavior, obsessive-compulsive neurotics, "adolescent reactions" in whom it was felt there was no serious pathology, incipient psychotics and overt psychotics as well as a group selected without regard to diagnosis. Comparisons of the scholastic averages of these groups were made with

averages of a randomly selected group of students attending the university during the same period of time.

The obsessive-compulsive neurotics did much better academically than did the random sample of the total student population with this difference statistically significant at better than the 1% level of confidence. The intelligence of these two groups was compared without finding significant difference between them. Comparison of the academic performance of the other experimental groups with the total student sample did not reveal any significant differences in scholastic achievement.

With the exception of the obsessive-compulsive group the findings tend to confirm the generalization derived from laboratory experiments that intellectual performance seems to be little affected by certain types psychopathology. The personality characteristics of the obsessive-compulsive individual appear to be syntonic with intellectual achievement of an academic type and the findings are thus compatible with clinical observation. The subjective complaints of impaired intellectual functioning by members of the other groups are not borne out by objective measurement of their academic accomplishments. Theories are presented relating to sparing of some ego functions in emotional illness to explain these findings but their relative lack of usefulness points out the need for further objective study of the ways in which ego functions are affected by psychopathology.

References

1. Fey, E. T. The performance of young schizophrenics and young normals in the Wisconsin Card Sorting Test. *J. consult. Psychol.*, 1951, 15, 311-319.
2. Greenfield, N. S. Some cognitive aspects of a personality dimension; neurosis and problem-solving behavior. Unpublished Ph.D. dissertation, Univ. of California, 1953.
3. Greenfield, N. S. Neurosis and problem-solving behavior. *J. clin. Psych.*, 1957, 13, 377-379.
4. Hartman, A. Ego psychology and the problem of adaptation. In Rapaport, D., (Ed.) *Organization and pathology of thought.* New York: Columbia Univ. Press, 1951.
5. Hunt, J. McV. and Cofer, C. N. Mental deficit. In Hunt, J. McV. (Ed.), *Personality and the behavior disorders.* Vol. 2. New York: Ronald Press, 1944.
6. Lazarus, R. S., Deese, J. and Osler, Sonia F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, 49, 293-317.
7. Matarazzo, J. D., Ulett, .G A., Guze, S. B. and Saslow, G. The relationship between anxiety level and certain measures of intelligence. *J. consult. Psychol.*, 1954, 17, 201-205.
8. Rapaport, D. The autonomy of the ego. *Bull. Menn. Clin.*, 1951, 15, 113-123.

SPONTANEOUS FLUCTUATIONS IN THE SEVERITY OF ILLNESS OF HOSPITALIZED LOBOTOMIZED PATIENTS[1]

THOMAS E. HANLON, CAROLYN S. SHEETS AND ALBERT A. KURLAND

Spring Grove State Hospital, Baltimore, Maryland

Problem

The present study was undertaken to answer the following questions relative to the observable behavior of lobotomized patients. Is there a significant variability over a prolonged period of time in the severity of the mental illnesses of hospitalized lobotomized patients who receive no therapy other than routine custodial care? When fluctuation is evidenced in the psychiatric status of a patient, to what personality components may we attribute this fluctuation? Do those lobotomized patients

[1]It is with a great deal of appreciation that acknowledgement is made of the authors' fruitful discussions with Dr. Maurice Lorr of the Veterans Administration and Dr. James P. O'Connor of the Catholic University of America concerning this project. It is also with a great deal of appreciation that acknowledgments are made to the Friends of Psychiatric Research, Inc., and the Women's Auxiliary of Spring Grove State Hospital, Inc., for their financial and volunteer assistance. A final note of appreciation is due to Dr. Isadore Tuerk, Superintendent of Spring Grove State Hospital, for his encouragement and interest in this project.

who show high fluctuation in behavioral ratings have certain identifying characteristics in common, and how do these compare with the characteristics of those who show low fluctuation?

Method

In a previous study by Kurland, *et al.*[3] the status of 122 patients who had been lobotomized over a ten year period at a typical state hospital were reviewed. At the time of this present study there were 80 patients of this group remaining in the hospital, who had either never left the hospital after their lobotomies or who had been unable to maintain a satisfactory adjustment outside of the hospital setting for any appreciable amount of time. All were considered treatment failures. Thirty-one of these cases, who had been on no treatment of any kind for a period of at least three months, were chosen as subjects for this study. All of these 31 patients had lobotomies performed at least a year prior to the study and had been hospitalized for periods exceeding two years. The age range of the sample was from 29 to 72 years, with a mean age of 46 years and a standard deviation of ten years.

The 31 patients comprising the sample were evaluated at approximately monthly intervals over a six month period with the Multidimensional Scale for Rating Psychiatric Patients (MSRPP)[4], which includes both ward and interview behavior observations. During this period these patients remained in the same environmental setting and received only routine custodial care. For each patient the MSRPP provided a total morbidity score, separate factor scores for the 11 factors comprising the scale, and three nosological pattern scores: hebephrenic, catatonic and paranoid. The average inter-rater reliability on the MSRPP for the three raters involved was .89. The reliability coefficients were determined on an independent, related sampling.

Results and Discussion

The severity of illness for the observed population according to mean MSRPP morbidity scores for the six rating situations ranged from a low of 19.3 for the best adjusted individual to a high of 73.7 for the least adjusted, with a mean score of 43.7 for the group and a standard deviation of 15.4. The mean score value is such that a patient obtaining this score would be placed at the 80th percentile when compared to the normative population for the MSRPP, *i.e.*, 80 percent of typical psychiatric hospital patients in the normative group obtained scores indicative of less morbidity. Two other total scores on the MSRPP representative of degree of mental illness, the "normality" and "difference" scores (similarity and dissimilarity of total pattern from what might be expected of normal subjects), were also obtained for each subject. However, since the correlations between the means of these scores and the mean morbidity scores were .91 and .94 respectively, they were disregarded in favor of the latter. The group mean morbidity scores for the six evaluations, together with the *t* ratios and correlations for the five between-rating situations may be found in Table 1.

Table 1. The Significance of the Differences Between the Means and the Pearsonian r's for the Six Rating Situations

Measures	1	2	3	4	5	6
Mean	40.81	43.35	44.16	44.58	43.71	43.45
Diffs		2.54	.81	.42	.87	.26
t*		1.85	.82	.37	.64	.15
r**		.88	.95	.94	.91	.82

*None significant (2.04 needed for significance at the .05 level).
**All significant at the .001 level.

As can be seen, the morbidity scores for the group are relatively stable over the six rating situations. None of the mean differences between successive ratings are statistically significant, while all of the correlations fall beyond the .001 level of significance. In a correlation matrix of the fifteen possible intercorrelations between ratings, none was found to fall below .80.

Bartlett's test for homogeneity of variance was performed for the six rating situations prior to the application of analysis of variance techniques in the determination of total mean variability. A resulting chi square of 1.57 showed that the variances were within limits of random sampling from a population with common variance so that no transformation of scores was necessary. An analysis of variance method accounting for the use of the same subjects over several testings was used to determine the significance of inter-rating mean variability[2]. The results of this analysis are presented in Table 2. The F ratio for between-ratings is 1.54, which is below the value of 2.27 needed for significance at the .05 level. The hypothesis of random sampling of the six mean scores from a common population is therefore upheld, *i.e.*, there is no significant inter-rating variability in morbidity for the group over the six rating sessions. (On the other hand, the extreme heterogeneity of the group with regard to morbidity may be noted by the highly significant F ratio for the between-patients variability.) When these same analysis of variance techniques were applied to the 11 factor and three pattern scores, in no case was the inter-rating mean variability found to be significant.

TABLE 2. ANALYSIS OF VARIANCE OF THE SCORES OF THE SAME SUBJECTS RATED AT SIX DIFFERENT MONTHLY PERIODS BY THE MSRPP

Source of Variation	Sum of Squares	df	Mean Squares	F
Between Ratings	272.23	5	54.45	1.54
Between Patients	43943.38	30	1464.78	41.31
Residual (error)	5318.37	150	35.46	
Total	49533.98	185		

Although there was no marked group trend in mean variability over the six rating sessions the fact that there was considerable random individual variability was evident from a cursory analysis of the tabulated morbidity scores for each member of the group. A more detailed approach to individual variability was therefore undertaken in order to gain insight into the personality components associated with these individual fluctuations. An index of fluctuation from one rating to another was determined for each patient by averaging two rank order methods, one considering primarily the patient's own variability over the six testing periods and the other taking into account his relative standing among the total group differences for each of the five between-rating intervals, since the mean differences and their distributions varied with each interval. The rank order correlation between the two methods was .75.

Eight (approximately 27%) of the 31 patients were then identified as "high fluctuators" and eight as "low fluctuators" according to their mean fluctuation rankings. Mean comparisons were made for these extreme groups on the MSRPP total morbidity score, the 11 factors underlying this score and the three nosological pattern scores. So that the entire sample might be considered, rank-order correlations were also determined between these scores and the fluctuation ranks of all the 31 patients (Table 3).

From the significantly high rank-order correlation of .56 between total morbidity and fluctuation and the correspondingly high t value, it is apparent that those patients who have the greatest degree of mental illness also tend to show the greatest degree of variability in behavior from one rating situation to another. From this, it would seem that any experimentation on the evaluation of behavioral or personality change in psychiatric patients should consider the initial differentiating character-

TABLE 3. THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN HIGH AND LOW FLUCTUATORS ON THE MSRPP TOTAL, FACTOR AND PATTERN SCORES, AND THE RANK-ORDER CORRELATION BETWEEN FLUCTUATION AND THESE SCORES

Score	Mean High Fluct.	Mean Low Fluct.	Diffs. (M_H-M_L)	*t* ratio	Rho† with Fluct.
MSRPP Total Morbidity	55.00	28.75	26.25	4.13**	.56***
Paranoid Pattern	14.00	17.86	-3.86	2.14*	-.36*
Hebephrenic Pattern	13.62	7.17	6.45	2.40*	.36*
Catatonic Pattern	9.38	3.77	5.61	2.43*	.50***
A Depressed-manic	18.21	15.52	2.69	1.45	.28
B Compliance-resistiveness	7.96	5.75	2.21	3.20**	.45**
C Paranoid Projection	6.40	5.67	.73	.51	-.01
D Inactive-overactive	14.02	13.94	.08	.09	-.08
E Melancholy agitation	16.54	14.06	2.48	2.67*	.32
F Perceptual distortion	12.60	8.48	4.12	2.33*	.42*
G Motor Disturbances	14.79	11.63	3.16	4.39***	.47**
H Submissiveness-belligerence	13.98	11.44	2.54	1.33	.25
I Withdrawal††	8.67	12.40	-3.73	4.34***	-.49**
J Self Depreciatory-grandiose	9.86	9.60	.26	.29	.18
K Conceptual disorganization	18.90	13.40	5.50	3.16**	.50***

*Significant at the .05 level
**Significant at the .01 level
***Significant at the .001 level
††Lower score indicative of greater withdrawal.
†Using $\frac{1.04\ (1-p^2)}{\sqrt{N-1}}$ as the standard error in determining significance, N = 31.

istic of the severity of the mental illness itself, since the significance of a purported change is indirectly dependent on the initial status of the patient at the outset. Precautionary measures to control this variable are especially important when either single cases or very small samples are involved.

From Table 3, it also may be seen that degree of fluctuation is related most highly to those factors characteristic of psychotics, specifically those schizophrenic processes of withdrawal (Factor I, and aspects of the catatonic pattern) and personality disorganization as evidenced in thinking, perception, and motor coordination (Factors F, K, and G and the general pattern of hebephrenia). However, there is a negative relationship found between fluctuation and the schizophrenic processes of paranoid reorganization and stability as evidenced on the paranoid pattern score. Of those factors of temperament which are observable in nonpsychotics, as well as psychotics (B, D, and H) only the resistive end of the bipolar Factor B, compliance-resistiveness, is significantly related to the amount of variability from one rating to another. Both of the affective factors (the excitement end of the bipolar Factor A and melancholy agitation, E) show slight, but not statistically significant, relationships to fluctuation. Only Factor E significantly differentiates the high and low fluctuating groups. The underlying component of this factor is depressive agitation. The other scores which significantly differentiate the two groups are the total morbidity score, the three nosological patterns, and Factors B, F, G, I, and K.

From the relationships described above, the clinical pictures of lobotomized patients with high and low behavioral fluctuations may be differentiated along the following lines. A patient whose hospital behavior is found to be erratic and unpredictable is likely to be one whose total personality makeup has become so disorganized as a result of his illness that there is an extreme amount of distortion in thinking and perception and a corresponding, generalized disturbance in motor control. Moreover, there is likely to be disorientation in any one, or all, of the three spheres. Withdrawal tendencies tend to be such that there are little or no interpersonal relationships maintained; contacts with others in the hospital are usually of a resistive and uncooperative nature. Finally, the affective state of such a patient is likely to be characterized by an hypertension stemming from underlying depressive elements.

The less disorganized patient tends to maintain a stationary clinical course and has been identified as a "low-fluctuator" in this study. On the whole, he has achieved a certain amount of reorganization and stability in his thinking and exhibits a minimal amount of psychotic, regressive behavior. The orientation of such a patient tends to be maintained, and comparatively adequate interpersonal relationships allow him to enjoy a fairly harmonious hospital adjustment. It is of interest to note the obvious similarities in stability of behavior between our typical low fluctuator and the S-3 schizophrenic reaction type identified by Beck[1] and elaborated upon by Molish *et al.*[5] in their theoretical study of the treatment failure.

Summary and Conclusions

The behavior of 31 hospitalized lobotomized schizophrenics was objectively measured at monthly intervals for a period of six months by means of the MSRPP. During this period the patients remained in their initial ward settings and received only routine custodial care. Inter-rating mean variability was determined for the MSRPP's total morbidity score, 11 factor scores, and three nosological pattern scores. The amount of individual variability was determined for each patient and the relationship between this variable and MSRPP scores studied. From an analysis of the results obtained the following conclusions are drawn:

1. The degree of mental illness for a group of schizophrenic patients, still hospitalized after being subjected to a variety of psychiatric treatments, including psychosurgery, remains relatively stable over a long term period. Although no group trends in behavior are discernible, there is much individual random variation to be noted.

2. Individual random fluctuation in hospital behavioral status for this group seems to be related to the severity of mental illness, with greater severity there is more variability. Future experimentation on the evaluation of change in the behavior of psychiatric patients needs to consider this as a differentiating characteristic in sample populations.

3. Degree of fluctuation is related most highly to the schizophrenic processes of withdrawal and personality disintegration as evidenced in thinking, perception and motor coordination. Resistiveness, as an aspect of temperament, and melancholy agitation are also related to behavioral fluctuations.

4. The restitutional factors evidenced in the schizophrenic processes of paranoid reorganization and stability are negatively related to spontaneous, random changes in the status of certain psychiatric patients.

References

1. Beck, S. J. *The Six Schizophrenias.* New York: American Orthopsychiatric Assoc., Inc., 1954.
2. Edwards, A. L. *Experimental Design in Psychological Research.* New York: Rinehart, 1956.
3. Kurland, A. A., Williams, T. G., and Hanlon, T. E. The response of chronically hospitalized, lobotomized patients to treatment with chlorpromazine and reserpine. In press.
4. Lorr, M. *Multidimensional Scale for Rating Psychiatric Patients 1: Hospital Form.* Veterans Administration Technical Bulletin TB10-507, Washington, D. C. 1953.
5. Molish, H. B., Hanlon, T. E., Kurland, A. A. A prognostic indicator of treatment failure in schizophrenia. *Arch. Neurol. Psychiat.*, 1957, 78, 177-193.

DIFFERENCES IN SOME PERSONAL CHARACTERISTICS OF "RIGID" AND "FLEXIBLE" INDIVIDUALS

K. WARNER SCHAIE

University of Nebraska[1]

PROBLEM

Much has been made in recent clinical discussions of the "Rigidity-Flexibility" dimension as an important behavior attribute. A great deal of research has been concerned with defining the single or multi-dimensional nature of this concept and a variety of tests have been constructed to provide operations for its measurement.[2] Several definitions have been proposed for specific or more general aspects of classes of behavior which are all related to difficulty or ease in adjusting to change in stimulus patterns and whose psychological characteristics are expressed in terms of a central concept of rigid and flexible behavior. In an earlier paper the following definition, which is believed to summarize most of the concepts advanced, was found to be useful. It defined behavioral rigidity as:

> . . . a tendency to perseverate and resist conceptual change, to resist the acquisition of new patterns of behavior, and to refuse to relinquish old and established patterns. (4)

This definition may seem rather broad, but a set of behavioral consequences can readily be deduced which may then be made amenable to observational test. One may deduce for example, that a rigid person dislikes to change patterns of interpersonal relationships; has difficulty in changing from one activity to another; is less efficient in activities requiring alternation and continous readjustment of motor activities, etc.

It was shown elsewhere that a battery of tests could be constructed to cover a variety of operations subsumed under the above definition. Subsequent to a series of factor analyses, it was also possible to show that the rigidity-flexibility concept can best be described by a three-dimensional system. In this sytem a "motor-cognitive rigidity" factor was seen to be a measure of effective adjustment to shifts in familiar patterns and to continuously changing situational demands; a "personality-perceptual rigidity" factor seemed to measure changes in cognitive and environmental patterns and new situations; and a "psychomotor speed" factor was considered to be related to the individual's functional efficiency in coping with familiar situations requiring rapid response and quick thinking. (4)

An attempt was made to relate scores on these factors to other behavior variables and it was found that a linear combination of the three factor scores was most successful in discriminating between the "rigid" and "flexible" groups. (5) The present paper is concerned therefore, with reporting some of the differences in the personal characteristics of subjects who were extreme high and low scorers on the composite rigidity measure.

PROCEDURE

Schaie's Test of Behavioral Rigidity (TBR) (4, 5) was administered to a sample of 500 subjects. The Ss had been selected by a stratified-random sampling procedure from a larger sampling universe and their characteristics (reported in (6)) suggest that they are probably quite representative of the upper 75% of the American socio-economic continuum. Ss ranged in age from 20 to 70 years and were divided into ten five-year interval sub-groups containing 25 men and 25 women each. In addition to a variety of psychological tests, Ss also filled out a schedule giving some of their personal characteristics. After these data had been punched on IBM cards, all age

[1]This paper is based on part of a dissertation submitted in partial satisfaction of requirements for the Ph.D. degree at University of Washington, Seattle. The author is indebted to Professors Charles R. Strother and Paul Horst for their advice and to the staff of Group Health Cooperative of Puget Sound (Seattle) for their assistance in obtaining the sample of subjects.

[2]Surveys of the literature on different aspects of the concept of "rigidity" may be found in papers by Werner (7) and by Levine (3); see also (6).

groups were separately rank-ordered on the Composite Rigidity score. The ten most rigid and the ten most flexible subjects in each age group were then selected for further study. Matching the extreme rigid and flexible individuals by age groups was required by the fact that rigidity-flexibility, just as many other behavior characteristics, shows significant age changes.[6] This method also provides more adequate sampling and permits wider generalization since the major portion of the adult age range is considered.

Results

Means for the extreme scorers were computed for measures of family status: age at first marriage, age the first child was born, and number of children. Measures of economic status included years of education, income level and occupational level. Income level was assessed on a six-point scale with $2000 intervals ranging from below $2000 to above $10,000. Occupational level was assessed using the major headings of the United States Census on a 10-point scale ranging from unskilled (0) to professional (9). Measures of mobility included frequency of change of residence, frequency of change of employment, and frequency of change of job classification. Self-ratings of happiness and success used a 5-point scale ranging from very happy or very successful (a score of 1) to very unhappy or very unsuccessful (a score of 5). A shortened version of the Social Responsibility scale constructed by Gough, McCloskey and Meehl[2] was also used. Table 1 gives the means for the rigid and flexible groups on all these variables.

Table 1. Means for the Personal Characteristics of the Extreme "Rigid" and "Flexible" Groups

Variables	Rigid (N = 100)*	Flexible (N = 100)	Difference
Family status:			
Age at marriage	24.14	25.20	1.06
Age at birth of first child	25.96	27.47	1.51
Number of children	2.26	2.09	0.25
Economic status:			
Years of education	10.94	14.93	3.99**
Occupational level	4.84	7.03	2.19**
Income level	2.70	3.50	0.80**
Mobility:			
Change of quarters	0.77	1.27	0.50**
Change of employment	0.84	0.67	0.17
Change of type of work	0.29	0.25	0.04
Self-ratings:			
Happiness	2.07	1.76	0.31**
Success	2.56	2.23	0.33**
Social responsibility	28.42	31.54	3.12**

*Since some of our subjects were unmarried, the N for the measures of family status are as follows: Age at marriage, rigid = 76, flexible = 70; age at birth of first child, rigid = 90, flexible = 85; number of children, rigid = 90, flexible = 90.

**Significant at or beyond the 5% level of confidence.

The analysis of variance was next used to test the null hypothesis for differences between the two groups on each variable and a factorial design, using methods suggested by Edwards[1], was employed which allowed for tests of the significance of differences between age groups, between flexible and rigid sub-groups, and of the interaction between age and rigidity-flexibility. Table 2 summarizes the results of these analyses. The F ratios are evaluated with one and 190 degrees of freedom for the difference between rigid and flexible groups and with nine and 190 degrees of freedom for the differences between age groups and for the interaction.[3] No F ratio is given where the test would be trivial.

[3]The F ratios on the measures of family status were evaluated with the reduced sample sizes as indicated in the footnote to Table 1.

TABLE 2. THE ANALYSIS OF VARIANCE TESTING THE DIFFERENCES IN GENERAL CHARACTERISTICS BETWEEN RIGID AND FLEXIBLE INDIVIDUALS AND BETWEEN AGE GROUPS

Variable	Rigidity-flexibility MS	F	Age MS	F	Interaction MS	F	Indiv. Diff.
Family status:							
Age at marriage	47.02	2.46	88.64	4.64*	25.65	1.34	19.12
Age at birth of first child	83.19	2.84	71.93	2.46*	12.73		29.25
Number of children	1.25		5.40	2.26*	3.06	1.28	2.39
Economic status:							
Years of education	796.01	133.33**	19.67	3.29**	7.50	1.26	5.97
Income level	32.00	36.36**	7.68	8.73**	1.48	1.68	0.88
Occupational level	239.81	96.16**	6.32	2.51**	2.08		2.52
Mobility:							
Change of residence	12.50	7.23**	16.99	9.82**	9.39	5.43**	1.73
Change of employment	1.45		7.12	3.14**	1.31		2.27
Change of occupation	0.08		1.41	3.44**	0.25		0.41
Self-ratings:							
Happiness	4.80	6.96**	0.61		1.48	2.14*	0.69
Success	5.45	5.50*	0.63		0.62		0.99
Social responsibility	486.72	27.30**	19.04	1.07	22.63	1.27	17.83

*Significant at the 5% level
**Significant at the 1% level

An examination of the mean scores for the rigid and flexible individuals on those variables where significant differences were shown in the analysis of variance discloses that the flexible individuals exceed the rigid ones in the positive direction in all cases. The extremely flexible members of the sample have had on the average four more years of education than the rigid. The mean income of the flexible Ss is 0.8 scale points or approximately $1600 higher. The mean occupational level of the flexible Ss exceeds that of the rigid by two scale intervals, indicating that our average flexible S was occupied at the managerial or proprietary level, while the average rigid S belonged to the skilled labor group. The flexible members of the sample also described themselves as more successful and happier, and they scored higher on the scale of social responsibility suggesting that they were more conscious of and subscribing to community maintaining values.

No significant differences are noted on the measures of mobility, except frequency of change of quarters. Here it is interesting to note a significant interaction between rigidity-flexibility and age. The younger flexible Ss exceed the rigid ones on residential mobility but the older rigid Ss exceed the flexible ones. One might speculate that, while in youth frequent changes of residence may be a function of taking advantage of changing opportunities, it may in the older group reflect difficulty in adjusting to environmental changes, there being a greater tendency for the rigid to move whenever the characteristics of their residential area changes. Another explanation for this reversal might be found in the fact that the lower income level of the rigid may force change in living quarters.

There was some tendency, although statistically not significant, for flexible people to marry later and to be older at the birth of their first child than the rigid members of the sample. The analysis also confirms significant age changes on all personal characteristics except for the self-ratings of happiness and success and the measure of social responsibility.

SUMMARY

500 subjects, half men and half women, ranging in age from 20 to 70 years, were divided into ten five-year age intervals and rank-ordered in terms of a composite

rigidity score. The ten most flexible and ten most rigid individuals in each group were examined with respect to certain of their personal characteristics. Results of an analysis of variance gave significant differences in favor of the flexible group on the measures of: years of education, income level, occupational level, self-ratings of happiness, success and social responsibility. Significant differences and interaction were found between rigidity-flexibility and age on a measure of mobility (change of residence) with the young flexible and the old rigid being the more mobile groups.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Gough, H. G., McCloskey, H. and Meehl, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, 47, 73-80.
3. Levine, A. S. Perseveration, rigidity and persistence. *Psychol. Reports*, 1955, 1, Monogr. Suppl. No. 5, 107-125.
4. Schaie, K. W. A test of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1955, 51, 604-610.
5. Schaie, K. W. Some developmental concomitants of rigid behavior. Unpubl. Ph.D. dissertation., Univer. of Washington, 1956.
6. Schaie, K. W. Rigidity-flexibility and intelligence: a cross-sectional study of the adult life span from 20 to 70. *Psychological Monographs*, 1957, 72, Whole No. 8.
7. Werner, H. The concept of rigidity. *Psychol. Rev.*, 1946, 53, 42-53.

FURTHER STUDIES ON THE BENDER-GESTALT TEST AND THE DIGIT-SPAN TEST AS MEASURES OF RECALL[1]

ALEXANDER TOLOR

USAF Hospital, Parks Air Force Base, California

Problem

In a previous study[3] significant group differences in recall on the Bender-Gestalt Test and the Digit-Span test were found when patients suffering from organic brain injury were compared with patients having a variety of psychogenic disorders. It was observed further that these differences tend to be more pronounced on the Bender-Gestalt Test than on the Digit-Span test. The present investigation is a replication study which has as its main purpose the determination of whether the previously noted differences are maintained when a better controlled experimental design and new samples of functional and organic patients are used.

In the present investigation two different groups of functional disorders were selected, *i.e.*, a group containing only patients diagnosed as having schizophrenic reactions and a group consisting of patients having only character and behavior disorders. The performance of these diagnostically homogeneous samples was then compared with that of a sample of patients having some type of organic brain syndrome. The use of these two markedly different groups of functional disorders also makes it possible to evaluate the effects of severity of emotional disturbance on Bender and digit recall.

Procedure

All patients participating in this study had been referred for psychological testing at Parks Air Force Base Hospital. The 18 patients comprising the character and behavior disorder group were all males. Included were 8 individuals diagnosed as having schizoid personalities, 2 with paranoid personalities, 1 antisocial personality,

[1]The opinions and conclusions do not necessarily represent those of the Department of the Air Force.

1 inadequate personality, 4 emotional instability reactions, 1 passive-aggressive reaction, and 1 passive dependency reaction.

Among the 18 psychotic patients there were 16 males and 2 females. The following diagnoses were represented in the group: 8 schizophrenic reactions, paranoid type; 2 schizophrenic reactions, latent type; and 8 schizophrenic reactions, not elsewhere classified. The diagnosis of schizophrenia was not generally an obvious one since ordinarily only the more complex cases constituting a diagnostic challenge were seen on referral.

All the patients in the organic group were male patients. The most frequent single diagnosis found here was that of encephalopathy due to trauma of which there were 8 cases. The other diagnoses covered a variety of organic disorders with demonstrable physical etiology or associated structural changes in the brain. In this category the most obvious cases of organicity were not generally seen but more frequently only those patients were encountered who represented a diagnostic problem at the beginning. It is precisely these less obvious cases who offer the clinician the opportunity of making his greatest contribution for diagnostic differentiation.

Since in the earlier study it had been demonstrated that both measures of recall are substantially correlated with intelligence, the present experimental design called for the matching for intelligence of each pair of subjects so that a strict controlling for this variable was accomplished. The subjects were also matched on the age variable as far as possible. Table 1 presents the essential characteristics of the three

TABLE 1. COMPOSITION OF THE THREE GROUPS

Item	Groups		
	Character & Behavior Disorder	Schizophrenic	Organic
N	18	18	18
Age			
Range	19–40	19–41	18–40
Mean	25.4	25.8	28.0
WAIS Full Scale IQ			
Range	77–124	78–123	77–125
Mean	99.5	99.4	98.8

groups of patients. The success of efforts to equate the groups is indicated by the fact that none of the group mean differences on these variables approached significance. It should be noted that the groups were formed without prior knowledge on the part of the examiner of the patients' performance on tests of recall.

The procedure utilized in obtaining Bender and digit recall scores was identical to that used in collecting data in the previous study [3]. Essentially this involved the individual testing of all subjects who were routinely referred for psychological testing. After the Bender-Gestalt Test was administered in a standard way [1], each patient was instructed to reproduce from memory as many designs as he could. In addition, patients were also requested to reproduce any part of the designs that they could recall. The total number of whole and part, correct and incorrect designs drawn constituted the overall Bender recall score.

Instead of using the Wechsler-Bellevue Scale, Form 1, as had been done formerly, in the present experiment the Wechsler Adult Intelligence Scale was administered. Inasmuch as the Digit-Span test is identical on the two scales, there was no difference in the manner in which the four digit recall scores, namely, total digits, digits forward, digits backward, and difference between the Digit-Span weighted score and the patient's average WAIS score, were obtained.

RESULTS

The major findings are contained in Table 2 presenting a comparison of the three groups of patients on five selected recall measures.

TABLE 2. SUMMARY OF BENDER-GESTALT TEST AND DIGIT-SPAN TEST RECALL SCORES FOR THREE DIFFERENT GROUPS (N = 18) OF PATIENTS

Recall Measure	Character & Behavior Disorder		Organic Brain Disease		Schizophrenic Reactions	
	Mean	SD	Mean	SD	Mean	SD
Bender Figures	6.3	1.28	4.9	2.01	5.8	1.53
Total Digits	11.0	1.59	9.9	2.04	11.7	2.45
Digits Forward	6.4	1.20	5.6	.95	6.9	1.44
Digits Backward	4.6	1.08	4.3	1.20	5.3	1.25
Difference Between Digit-Span and Mean WAIS Score	.5	2.11	−1.4	2.62	.7	2.94

Table 3 shows the level of significance attained when critical ratios are computed in comparing each group's performance on the five measures of recall with every other group. Since the groups were matched for intelligence, the following *r* were introduced into the formula for the standard error of the difference between means: .50 between intelligence and Bender recall, .70 between intelligence and total digit recall, .60 between intelligence and digits recalled forward, .65 between intelligence and digits recalled backward, and .70 between intelligence and the difference between the Digit-Span test and the mean WAIS score. These correlations were taken from the previously performed study.

TABLE 3. CRITICAL RATIOS AND LEVELS OF SIGNIFICANCE IN COMPARING THREE GROUPS OF PATIENTS ON SEVERAL MEASURES OF RECALL

Measures of Recall	Character & Behavior *vs.* Organics	Schizophrenics *vs.* Organics	Character & Behavior *vs.* Schizophrenics
Bender Figures	3.26**	2.05*	1.44
Total Digits	3.06**	4.14**	1.63
Digits Forward	3.33**	4.66**	1.71
Digits Backward	1.30	4.03**	2.74**
Difference Between Digit-Span and Mean WAIS	4.13**	4.04**	1.63

*Significant at the .05 level or better
**Significant at the .01 level or better

It will be observed from Table 3 that with the exception of number of digits recalled backward, all of the recall measures selected reveal a significantly superior performance of the patients belonging to the character and behavior pattern group as compared to the organic group. In comparing the organic with the schizophrenic patients, all of the recall scores for the schizophrenics exceed to a significant degree the recall scores obtained by the patients with organic brain injuries. On the other hand, with the exception of the significantly greater recall of digits backward for schizophrenic as compared to character and behavior disorder patients, none of the differences between these two psychogenic groups are significant.

Further comparisons were made between groups using the Median Test. In applying this method to test the hypothesis that the population of organic patients has a smaller median in total number of Bender figures recalled than the population of character and behavior disorder patients, confirmation is found for this hypothesis. The chi square of 2.76 is significant at the .05 level using a one-tailed test. A similar test with respect to total number of digits recalled indicates that the organic patients tend to have a smaller median than patients with character and behavior disorders (chi square of 2.76 is significant at the .05 level using a one-tailed test). The Median Test failed to support the hypothesis that a greater number of organic cases will fall below the combined median for organic and schizophrenic patients with reference to total number of Bender designs recalled. However, support was found for the hypothesis that a greater number of organic patients will fall beneath the combined

median for the two groups in total number of digits recalled. The chi square of 4.08 yielded a P of .025 using a one-sided test of significance. In each of these non-parametric analyses the Yates' correction was applied.

The frequency with which a mixed group of psychogenic patients and a sample of organic patients recall the various Bender designs has been reported by Tolor[3]. An attempt is made here to report similar data on more homogeneous groups in order to determine the stability of differences in the frequency of design recall. Table 4 presents a breakdown of the proportion of each group recalling each one of the nine Bender designs as well as the rank order of recall.

Although the three groups do differ somewhat in the ease with which patients recall the various designs, in general Designs 3 and 4 appear to be among the most difficult whereas Figures 8 and 6 are usually found among the easiest. These results are similar to those obtained in the above-mentioned study as well as those reported by Reznikoff and Olin[2]. In contrast to the later investigation, however, the present experiment shows Design 7 to be one of the easiest for recall.

Further evidence for the high degree of similarity in the Bender recall pattern is derived from the high rank-difference correlations between groups. The highest correlation (.90) is found between the organic and character and behavior disorder groups whereas the lowest (.75) is obtained between the schizophrenic and organic groups. The correlation between schizophrenic and character-behavior disorder recall is .85.

Discussion

The results of this study provide further support for the hypothesis that there are significant differences in immediate recall, as measured by the Bender-Gestalt Test and the Digit-Span test, between organic and psychogenic patients. As was noted previously, these group differences are certainly insufficient in magnitude to be of diagnostic value for individual cases; however, they do consistently demonstrate that memory impairment tends to be far more severe if secondary to an organic brain injury than if it is a consequence of a functional disorder. These results are even more impressive when one considers the strict matching of subjects on the intelligence variable and the close matching for age, both of which might otherwise have resulted in obscure or inconclusive results.

In addition, the results indicate that with the exception of one measure of recall, namely, the number of digits repeated backward, the severity of the functional disorder has no effect whatsoever on the basic finding that the immediate memory of psychogenic patients tends to be superior to that of organic patients. Thus, patients who have developed some type of character defect resemble much more closely in their ability to recall test material patients who are psychotic than they do individuals with some kind of organic brain syndrome. Reznikoff and Olin[2] also noted the superior performance of schizophrenic patients as compared to organics, but only on one of their three measures of Bender recall.

Table 4. Proportion of Subjects in the Three Groups Recalling Each of the Nine Bender Designs and Rank Order of Recall

Design	Organics		Schizophrenics		Character & Behavior Disorders	
	Proportion	Rank	Proportion	Rank	Proportion	Rank
A	.44	6.5	.56	6	.56	7
1	.44	6.5	.94	2	.84	4
2	.56	4.5	.67	5	.72	5
3	.22	9	.22	9	.50	8
4	.28	8	.28	8	.28	9
5	.56	4.5	.50	7	.67	6
6	.78	2	.84	3.5	.94	1
7	.67	3	.84	3.5	.88	2.5
8	1.00	1	1.00	1	.88	2.5

Inasmuch as the character and behavior pattern group did not differ significantly from the organic group in number of digits recalled backward, whereas both psychogenic groups differed from the organic group in Bender figures recalled, there is partial but inconclusive support in this study for the impression that the Bender recall is a better measure for discrimination than the Digit-Span test. The poor performance of the character and behavior disorder group in recalling digits backward might be a function of these patients' relative intolerance to strong situational stress. This stress, it would be reasonable to assume, is greater for recalling digits backward than for repeating them forward.

Summary

Three groups of patients, a schizophrenic group, a character and behavior disorder group, and an organic brain injury group were compared for their performance on several measures of recall which had been derived from the Bender-Gestalt Test and the Digit-Span test. All patients had been matched for intelligence and age. The results were in essential agreement with previous findings in demonstrating that the performance of psychogenic patients tends to be superior to that of organic patients on both tests. It was shown further that the severity of the functional disorder bears little relationship to the differences between organic and non-organic groups and that the order of difficulty in recalling the nine Bender designs tends to be similar for all samples tested.

References

1. Bender, Lauretta. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. Res. Monogr.*, 1938, No. 3.
2. Reznikoff, M. and Olin, T. D. Recall of the Bender-Gestalt designs by organic and schizophrenic patients; a comparative study. *J. clin. Psychol.*, 1957, *13*, 183-186.
3. Tolor, A. A comparison of the Bender-Gestalt Test and the Digit-Span test as measures of recall. *J. consult. Psychol.*, 1956, *20*, 305-309.

AGE, INTELLIGENCE, APHASIA AND THE SPIRAL AFTER-EFFECT IN AN EPILEPTIC POPULATION

BERNARD S. AARONSON

New Castle State Hospital, New Castle, Indiana

Problem

The ability to perceive the spiral aftereffect illusion has been proposed by Price and Deabler[7] as a means of distinguishing between individuals with and without brain damage. A number of other investigators[1, 2, 6] have shown that the procedure has diagnostic utility, although most have not reported the striking results noted in the original article. The present study is an attempt to examine this phenomenon as it appears within the context of an epileptic population with an high incidence of mental deficiency. By examining the relationship between performance on this task and such variables as age, intelligence, severity of anomia, extent of disorder in sensory perception, and laterality of sensory disorders, an attempt is made to isolate some of the factors entering into performance on this task.

Procedures

The spiral aftereffect was administered to 65 consecutive patients seen in the psychology department for any diagnostic purpose. Patients not in sufficient con-

tact to take the test were excluded. The sample consisted of 35 males and 30 females who ranged in age from 8 to 70. The mean age was 31.3 with an S.D. of 17.1 years.

The stimulus material consisted of two six-inch white cardboard disks with a 920° Archimedes spiral painted on each with black paint. The spirals pointed in opposite directions so that they gave opposite effects of rotation when placed on a 78 rpm phonograph turntable. The method of administration and scoring followed Gallese[1] and minimized the effect of verbal factors. Each stimulus card was presented twice in rotating order and a point was scored every time the aftereffect was correctly reported. An aftereffect of contraction was considered "correct" for the stimulus which seemed to expand during presentation, and one of expansion was considered "correct" for the stimulus which seemed to contract. The range of possible scores was from 0 to 4.

The distribution of scores was obtained and the percentage of cases obtaining scores of two or less compared with that reported by Gallese. For the remainder of the comparisons, the sample was dichotomized between those perceiving the proper aftereffect on all presentations and those failing on any presentation. This was done as it was felt that in this population where organicity is the presumption, the use of a lower cutting score would tend to obscure some of the true variance.

The sample was divided into three equal groups on the basis of age, and the interaction of age and spiral performance tested by chi square. As a measure of the relationship between these two variables when viewed against a normal surface, a biserial correlation was computed.

The Wechsler-Bellevue Intelligence Scale, Form I was administered to 40 patients in this sample. This group consisted of 22 males and 18 females with a mean age of 32.6 and an S.D. of 15.2 years. Their verbal, performance, and full scale IQs were compared with their spiral performance by chi square and biserial correlation. The chi square values were corrected by Yates' correction for continuity.

An aphasia screening test devised by Reitan[4] was administered to 50 patients in this sample. This group consisted of 27 males and 23 females, with a mean age of 30.4 and an S.D. of 17.4 years. Performance on the spiral was compared by chi square, which was corrected for continuity, with the ability to identify orally the drawings of three "concrete" and three "abstract" objects. The concrete objects were a baby, a fork, and a key. The abstract objects were a square, a cross, and a triangle.

This test also includes a set of procedures to test for impairment of visual, auditory, and tactual receptive function. These procedures measure failures of perception of stimulation through these modalities and suppression of stimulation to one side when the modality is stimulated bilaterally. Forty seven patients produced usable records with regard to all of these modalities. Some patients had to be eliminated because of confusion as to which side was being stimulated in a given sensory modality and some because of chronic conditions which made it difficult to test in that area. These included chronic deafness, very poor visual acuity, and the absence of hands. Performance on the spiral was compared by chi square with the number of different sensory areas in which impairment was noted. Since the absolute number of those showing impairment in any given sensory avenue was too small to meet the assumptions of chi square, the hypothesis that impairment would go along with lowered spiral performance was tested by Yule's coefficient of association, Q[5]. Finally, the 22 cases showing a preponderance of right-sided sensory involvements were compared with regard to their spiral performance by chi square, which was corrected for continuity, and by Q with the 12 who showed left-sided involvements.

Results

As shown in Table 1, the distribution of spiral scores tends to be U-shaped. In the present sample, 44% obtained scores of two or less, which is a greater number than that reported by Gallese[1] for his mixed epileptic and alcoholic groups and comparable with the results reported with the remainder of his organics.

TABLE 1. DISTRIBUTION OF SPIRAL SCORES IN THE SAMPLE AND THE PROPORTION OBTAINING EACH SCORE

Score	N	Proportion
4	28	43
3	9	14
2	7	11
1	7	11
0	14	21

When performance on the spiral is compared with age, no relationships are noted. A chi square of 0.86 is obtained, which is not significant. Inspection of the actual scores, however, does suggest a trend for older individuals to obtain poorer scores. A biserial correlation for these data is −.21, which is not significant.

Comparison of spiral performance with Wechsler scores yields significant associations with verbal, performance and full scale IQs. The chi square values for spiral performance with these scores are 6.40, 10.00, and 6.40 respectively, which are significant at the .02 and .01 levels. Biserial correlations calculated for these same data are .55, .66, and .63 respectively, of which the first is significant at the .01 level and the latter two at the .001 level.

Only four patients in the sample failed to name all three drawings of concrete objects and none of these reported the aftereffect correctly on all four exposures. In comparing spiral performance with ability to name geometrical shapes, one error in naming was allowed because of the low educational level of the group. The distribution of naming scores was dichotomized between one and two errors. The chi square obtained in this comparison is 10.15, significant at beyond the .01 level.

When spiral performance is compared with the number of sensory modalities in which disorders of perception occur, a chi square of 3.81 is obtained, significant at the .06 level. The coefficients of association between spiral performance and specific sensory impairments are shown in Table 2. Poor spiral performance seems highly associated with impairment of visual function and to a lesser extent with auditory impairment.

TABLE 2. COEFFICIENTS OF ASSOCIATION BETWEEN SENSORY IMPAIRMENT AND SPIRAL AFTEREFFECT FAILURES

Area of Impairment	N		Q
	Impaired	Nonimpaired	
Auditory	11	37	.48*
Visual	11	38	.70**
Tactual	21	28	.30

*p is at least .05
**p is at least .001

Comparison of the relation of laterality of sensory impairment with spiral performance yields a chi square of 1.16, which is not significant. Since a trend for lower scores to be associated with right-sided sensory impairment was noted, Yule's coefficient of association was calculated and a Q value of .49 obtained, significant at the .08 level.

The data support the validity of the spiral aftereffect as a measure of brain damage. The agreement noted with Gallese's results [1], the tendency for performance to drop with IQ, anomia, and extent of sensory impairment all suggest that performance on this task tends to decline as brain damage becomes more severe.

While the greater association with the Performance IQ on the Wechsler might tend to suggest that the test is tapping a right-sided brain function [8], the strong association with anomia tends to implicate the left hemisphere. This last seems to

be buttressed by the trend for individuals with right-sided sensory impairment to get poorer spiral scores. If the pattern of association between the spiral and visual, auditory, and tactual impairment is taken into account, the data suggest a primary involvement of the visual pathways in the temporal area. An involvement impinging on the visual pathways in this region would also tend to interrupt auditory fibers. Involvements posterior to this might show more relation with tactual impairment and less with auditory.

The association of lowered spiral performance with anomia raises the question as to what extent impaired performance is a function of being unable to perceive the spiral aftereffect and to what extent it is a function of being unable to express verbally changes in an ambiguous stimulus. Goldstein and Scheerer[3] have previously pointed out the difficulty that organics have in conceptualizing ambiguous stimuli. Insofar as left temporal lobe involvement tends to be associated with impairment of language function and in these data there is a suggestion of an association between both temporal lobe involvement and left hemisphere involvement and spiral performance, the extent to which performance on this task is a function of verbal factor requires further evaluation.

Summary

The spiral aftereffect was administered to 65 epileptic patients of whom 40 had also taken the Wechsler-Bellevue, Form I, and 50 had taken an aphasia screening test. The percentage of individuals making two or more errors agreed with that found by Gallese for organics. No relation to aging was found. Poorer spiral performance seemed associated with lower Wechsler scores, anomia, the number of different sensory avenues impaired, visual and auditory impairment, and right-sided sensory impairment. The data suggest that the test is most sensitive to temporal lobe involvement and may be most sensitive to left cerebral hemisphere involvement. The question is raised as to whether lowered performance on the spiral arises from inability to perceive the aftereffect or inability to verbalize changes in an ambiguous stimulus.

References

1. Gallese, A. J., Jr. Spiral after-effect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-258.
2. Garrett, E. S., Price, A. C. and Deabler, H. L. Diagnostic testing for cortical brain impairment. *Arch. Neurol. Psychiat.*, 1957, 77, 223-225.
3. Goldstein, K. and Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).
4. Heimburger, R. F. and Reitan, R. M. *Testing for aphasia and related disorders.* Indianapolis, Ind.: Department of Surgery (Section of Neurological Surgery), Indiana University Medical Center, 1957.
5. Moroney, M. J. *Facts from figures.* Harmondsworth, Middlesex, England: Penguin, 1951.
6. Page, H. A., Rakita, G., Kaplan, H. K. and Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89-91.
7. Price, A. C. and Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, 19, 299-302.
8. Reitan, R. M. Certain differential effects of left and right cerebral lesions in human adults. *J. comp. physiol. Psychol.*, 1955, 48, 474-477.

WECHSLER BELLEVUE CHANGES OVER TIME

STEWART G. ARMITAGE AND DAVID PEARL

Veterans Administration Hospital, Battle Creek, Michigan

Problem

Numerous unsuccessful attempts have been made to determine cognitive patternings on the Wechsler Bellevue Intelligence scale which could be utilized in the differential diagnosis of psychopathology. Many clinicians, however, continue to envisage varying aspects of intellectual functioning as being selectively affected by differing psychopathological processes. One such distinction is often made between the effects of schizophrenic processes and non-psychotic (non-brain damaged) maladjustments on cognition. While maladjusted non-psychotics may demonstrate inefficiencies in intellectual functioning, their performance is assumed to remain more stable than is that of persons undergoing a schizophrenic process. This concept of deficit in schizophrenia as an ongoing process can be detected in the distinction often made between the early and the more chronic schizophrenic patient. Some have held that in early schizophrenia, obvious intellectual impairment is not apparent, subtle qualitative evaluations of test performance being necessary to detect beginning thought deficits, whereas intellectual loss is more easily seen in the more chronic schizophrenic patient. A further distinction often is made between schizophrenic categories. Thus, the paranoid schizophrenic patient supposedly undergoes less loss than the other types.

The purpose of this study was to investigate the concept of process as applied to intellectual functioning in the instance of schizophrenic and non-schizophrenic, non-brain damaged subjects using the Wechsler-Bellevue Intelligence Scale, Form I, which was given to a group of 162 hospitalized schizophrenic and non-psychotic (non-brain injured) subjects upon their initial and again upon a following hospitalization. It was proposed to determine the stability of intellectual performance over time through the correlation of initial and subsequent Wechsler-Bellevue performance, and to evaluate the following two hypotheses: (a) that greater intellectual deficit will occur over time in schizophrenic subjects than in non-psychotic patients and (b) that paranoid schizophrenics undergo less intellectual deficit over time than other schizophrenic subjects, in this instance, those diagnosed as unclassified schizophrenics.

Procedure and Results

As previously indicated, 162 rehospitalized neuropsychiatric patients diagnosed similarly on both hospital admissions as either paranoid or unclassified schizophrenics or non-psychotic were retested with the Wechsler-Bellevue scale. Mean age for patients upon first hospitalization was 30.5 years (SD 7.5) and average length of time between hospitalizations was 1.5 years. Only the above subcategories of schizophrenia were utilized, since too few cases of the other types were encountered. Since none of these subjects had been hospitalized prior to their first admission, it may be assumed that they, particularly the schizophrenic subjects, were in the early or less chronic stages of their maladjustment at the time of initial hospitalization.

Data bearing on the question of stability was obtained through linear correlations of initial and readmission scores on Wechsler-Bellevue subtests, verbal, performance and full scale IQs, and are presented in Table 1. Although these cannot be considered as measures of reliability since an average of 18 months occurred between tests and clinical change may have resulted, they are comparable to Wechsler-Bellevue scale reliabilities found by other investigators [1, 2, 3, 4].

In order to determine intellectual changes between hospitalizations, elapsed time between test administrations was trichotomized on the basis of its distribution. Since ageing of subjects may affect Wechsler-Bellevue changes, it was controlled by considering test score changes for subjects of three age categories, 25 years or less, 26-32 years and 33 years and over. Evaluation of the two hypotheses was investi-

TABLE 1. WECHSLER-BELLEVUE INTELLIGENCE SCALE, FORM I TEST-RETEST CORRELATIONS

	Our Unclassified and Paranoid Schizophrenics and Misc. Non-Psychotic Diagnoses (N = 162)	Schizophrenics [2] (N = 53)	Miscellaneous Diagnoses [4] (N = 60)	Miscellaneous Schizophrenic Diagnoses [3] (N = 30)	Normals, Combined Average [1] (N = 158)
Information	.85	.94	.99	.89	.86
Comprehension	.65	.76	.44	.12	.74
Arithmetic	.68	.87	.68	.75	.62
Digit Span	.57	.80	.77	.62	.67
Similarities	.68	.86	.62	.38	.71
Vocabulary	.91	.90			.88
Picture Arrangement	.55	.73	.60	.54	.64
Picture Completion	.61	.71	.89	.32	.83
Object Assembly	.49	.66	.67	.31	.69
Block Design	.78	.71	.65	.71	.84
Digit Symbol	.72	.80	.74	.34	.80
Verbal Scale IQ	.84	.91	.89	.78	.84
Performance Scale IQ	.77	.83	.76	.52	.86
Full Scale IQ	.84	.87	.84	.54	.90

gated through analyses of variance by means of a triple classification factorial design in which the variables of diagnostic category, subjects' ages at initial testing and elapsed time between successive test performances were each varied three ways. Since the use of all 162 subjects did not provide equal or proportional representation for all cells of the factorial design, it was necessary to set up a ratio of four unclassified schizophrenic to two each of paranoid schizophrenic and non-psychotic subjects for each factorial cell involving age and test time elapsed. A total of 36 unclassified schizophrenic and 18 each of the other two diagnoses were randomly selected from the larger subject population for these analyses. The nature of the variance analyses is illustrated in Table 2.

TABLE 2. ANALYSIS OF VARIANCE DESIGN FOR WECHSLER-BELLEVUE SCALES AND SUBTESTS

Source	Degrees of Freedom
Diagnosis	2
Time Interval	2
Age	2
Diagnosis x Time Interval	4
Diagnosis x Age	4
Time Interval x Age	4
Diagnosis x Time Interval x Age	8
Error	45
Total	71

In general, results of these analyses fail to show any of these variables or their interactions as exercising a significant effect upon Wechsler-Bellevue score changes. Only two variables were found to be significantly related to change, elapsed time for Picture Arrangements and diagnosis for the Vocabulary subtest. In the first, subjects retested within nine months obtained improved scores, whereas no changes occurred for those with a longer interval period, a result which may be attributable to a practice effect. In the second, non-psychotic subjects' vocabulary improved, while no change occurred for the two schizophrenic groups. Two variable interactions were

significant, Age x Diagnosis in the Arithmetic subtest and Elapsed Time x Age in the Digit Symbol subtest. Since 98 tests of significance were made and only four significant relationships were found, these must be considered as probably due to chance, so that neither of the study's two hypotheses are supported by the data.

Discussion

Despite an average elapsed time of 18 months between tests, achievement on Information and Vocabulary subtests remained highly constant as did Verbal and Full Scale IQs and to a lesser extent, Performance Scale IQs and Block Design subtest scores. Table 1 indicates that test-retest correlations of this study are generally similar to those found by previous investigators. Considering the unreliability of many of the subtests, it is indeed surprising that there should be any substantial agreement. Such correspondence would suggest that any changes observed in test scores of the study's subjects may be attributable to test unreliability rather than to any clinical changes occurring between tests.

Summary

The stability of Wechsler-Bellevue Intelligence Scale, Form I, performance of 162 schizophrenic and non-psychotic patients was determined by correlating initial test achievement with test performance on a subsequent hospitalization. Deficit occurring in unclassified, paranoid schizophrenics and non-psychotic patients in the interval between tests was investigated by means of a factorial design with diagnosis, subjects' ages and intervening time between tests as variables. In general, test-retest correlations are similar to those found in reliability studies. No differences were found between the three diagnostic groups with respect to changes on various Wechsler-Bellevue measures.

References

1. Derner, G. F., Aborn, M. and Canter, A. H. Reliability of the Wechsler-Bellevue. *J. consult. Psychol.*, 1950, *14*, 172-179.
2. Hamister, R. C. Test-retest reliability of the Wechsler-Bellevue. *J. consult. Psychol.*, 1949, *13*, 39-44.
3. Rabin, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, *18*, 78-91.
4. Rabin, A. I. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, *31*, 231-239.

A COMPARISON OF THE BENDER GESTALT REPRODUCTIONS OF DELINQUENTS AND NON-DELINQUENTS[1]

EDWIN S. ZOLIK

Marquette University

Problem

The Bender Gestalt (B-G) Test has been little used with juvenile delinquents, a group in which emotional maladjustments may be assumed to play a conspicuous role[2]. Since the B-G assumes that the reproduction of visuo-motor gestalten is influenced and distorted by intellectual impairment and emotional maladjustments [1, 4, 5], we may hypothesize that when maturational and intellectual factors are held constant, the test performance of juvenile delinquents and non-delinquents will differ significantly.

---

[1]This study is based upon a thesis submitted by the writer to the Graduate School of Arts and Sciences of the Catholic University of America in partial fulfillment of the requirements for the M. A. degree in psychology. The writer is indebted to Professor William D. Commins and Dr. Helen Peixotto for their valuable assistance. We also wish to thank Dr. Raymond McCall for his critical reading of this manuscript.

Method

Forty-three adolescent delinquents were individually matched with a group of forty-three non-delinquents, residing in a high delinquency area, in the factors of age, Otis IQ, and absence of any type of motor defect. An initial attempt to equate the groups on the variable of education did not prove feasible. The mean age of the control group is 15.95 years, (s. d. = 0.92), whereas the delinquent group has a mean of 16.12 years, (s.d. = 0.90). The mean Otis IQ for the control group is 105.28 (s. d. = 8.90) and for the delinquent group the mean is 101.65, (s. d. = 8.39). The Pearson *r* between the ages of the two groups is .81 and between the IQ's is .61, both of which are significant at the .001 level.

Results

An analysis of the B-G figures, comparing the control and delinquent groups, was made with respect to both the sum total score and the individual figure scores employing Pascal and Suttell's[5] method. Table 1 gives the critical ratios obtained and their significance. In order to ascertain which scored aspects of each figure contributed to these significant differences, the performances on each figure were further analyzed and critical ratios computed.

Table 1. Comparison of Delinquent and Non-Delinquent Scores on the Bender-Gestalt Figures (N = 43)

Figure	Non-delinquent Mean	Non-delinquent SD	Delinquent Mean	Delinquent SD	CR	P
1	1.54	1.72	3.91	2.58	6.61	.001
2	2.19	2.07	3.51	2.79	2.46	.05
3	2.70	2.49	5.91	4.22	4.25	.001
4	1.63	1.60	3.12	2.40	3.35	.001
5	2.63	1.83	5.51	3.11	5.17	.001
6	2.05	2.19	4.58	3.53	3.95	.001
7	1.05	1.58	4.28	3.91	4.97	.001
8	1.09	1.54	3.67	4.18	1.35	——
Configuration	.91	1.90	1.84	2.82	1.36	——
Total	47.72	7.29	70.79	2.82	7.25	.001

"Asymmetry" is significant (.01 level) on the three figures (Figs. 3, 5, 6) where it is scored.

The variable "Dots, Dashes, Circles" is significant (.01 level) on two of the three figures (Figs. 3, 5) where it is a scored deviation. From Pascal and Suttell's results [5, p. 48] it appears that this is indicative of the possibility of either a maturational failure or of emotional components giving rise to regressive tendencies in our delinquent group.

"Tremor" is a significant indicator (Fig. 4: .05 level; Fig. 7: .001 level) on two of the four figures on which it is a scorable variable. This may be indicative of a higher degree of neuromuscular incoordination under conditions of tension and anxiety in the delinquent.

"Second attempt", a scorable deviation on all the figures, was significant (.05 level) on Fig. 4 only. Hutt[4] considers the curvilinear part of Fig. 4 to have dynamic implications concerning an individual's relationships with female figures. Since the "second attempt" occurred on this part of the figure in all the delinquent records, it appears that further research should be directed at substantiating Hutt's formulation.

"Wavy line" is significant on Fig. 1 (.05 level). "Dashes" is significant on Fig. 1 (.05 level). "Extra angles" is significant on Fig. 7 (.01 level).

Pearson *r*'s were computed between age and intelligence and the B-G scores for both the delinquent and non-delinquent groups. The correlation between IQ and the B-G score was –.13 for the delinquent group and –.16 for the non-delinquent group,

neither correlation reaching significance. Between chronological age and B-G score r is .04 for the delinquent group and .01 for the non-delinquents, neither of which is significant. The obtained coefficients are comparable to those reported by Pascal and Suttell [5, p. 21, 23], and it can be concluded that, after maturation, errors which appear in the reproduction of the B-G figures are not attributable to chronological age nor intelligence but to other psychological differences. Finally, a Pearson r of −.10 between the B-G scores of the matched pairs in both groups indicated that equating did not lead to a significant correlation.

To determine the validity of the dichotomy between delinquents and non-delinquents, a tetrachoric r was computed according to the Cosine-pi formula [3, p. 336] using the median B-G z score as the point for dichotomizing the two groups. The obtaining r_{tet} of $-.77 \pm .093$ is more than eight times its S. E. and highly significant. The efficiency of this dichotomization was improved by curtailing Pascal and Suttell's questionable range, from 50-72 to 50-60. An upper cut-off score one S. D. below the mean of the delinquent group, or a z score of 60, excludes 95.4% of the control group. Of the delinquents, 20.9% would fall in the questionable range but the number of delinquents correctly classified would increase from 32.6% to 69.8%. This increase in diagnostic prediction is significant at the .01 level.

Summary

The Pascal and Suttell scoring method of the B-G Test has proved valid for differentiating between delinquents and non-delinquents. The use of a different questionable range than that reported by Pascal and Suttell is suggested for enhancing the dichotomization of groups in diagnostic prediction.

References

1. Bender, L. *A visual motor gestalt test and its clinical use.* New York: Amer. Orthopsychiat. Assn., 1938.
2. Glueck, S. and Glueck, E. *Unraveling juvenile delinquency.* New York: Commonwealth Fund, 1950.
3. Guilford, J. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.
4. Hutt, M. Bender gestalt workshop. E. S. Zolik (Ed.), Unpublished Catholic University of America Workshop, 1950.
5. Pascal, G. and Suttell, Barbara. *The Bender Gestalt Test.* New York: Grune and Stratton, 1951.
6. Zolik, E. *The reproduction of visuo-motor gestalten by delinquent and nondelinquent children.* Unpublished M.A. thesis, The Catholic University of America, 1952.

PEAQ: A MEASURE OF PSYCHOPATHIC BEHAVIOR[1]

MORTIMER H. APPLEZWEIG, *Connecticut College*
ANDREW S. DIBNER, *Clark University*
RAYMOND L. OSBORNE, *New York City*

Introduction

Preu defines psychopathy as designating "... those individuals who have manifested considerable difficulty in social adjustment over a period of many years or throughout life, but who are not of defective intelligence nor suffering from structural disease of the brain or epilepsy, and whose difficulties in adjustment have not been manifested by the behavioral syndromes which are conventionally referred to as neuroses and psychoses [8, p. 923]." Despite a number of attempts to dismiss the category as a wastebasket classification, a useless conglomerate, a misclassification, etc. [7, 8], the terms *psychopathic personality, psychopathic inferior, psychopathic deviate, constitutional psychopathic state, sociopathic personality, anethopath,* etc., are

[1]Presented at the meetings of the Eastern Psychological Association in New York City, April 12, 1957.

still widely used today for the classification of persons with behavior and/or adjustment problems.

The present paper is a description of an attempt to psychometrically identify psychopathic behavior. The first form of the Personal Experience and Attitude Questionnaire[2] was developed some 13 years ago, during World War II, at an Army Air Force Regional Hospital in Colorado. The large number of referrals to an understaffed clinic made the use of a screening battery of tests mandatory. The unique problems of military adjustment made the identification of psychopathy particularly important. Dr. Osborne, then chief of neuropsychiatry, was responsible for the initial selection of items for the scale, and Dr. Applezweig for the initial standardization. The Inventory was found useful in clinical work, but before it could be properly evaluated, changes in military assignments led to its temporary abandonment, and it was not taken up again until 1949, by Drs. Applezweig and Dibner at the University of Michigan. Here a rewriting of the items and a shortening of the scale was undertaken on the basis of an item analysis of the earlier data and new data obtained on a large undergraduate male sample.

Brief Description of the PEAQ

The PEAQ is a true-false questionnaire, consisting of 150 items, 10 of which constitute a validity scale and do not contribute to the total score. The 140 items composing the psychopathy scale concern life experiences or attitudes in six general areas, as follows:

Criminalism: 25 items (*e.g.*, I go for easy money, whether it s right or wrong. Two or more of my friends have been in jail. Most policemen are crooked and will accept graft.)

Emotional instability: 20 items (*e.g.*, My childhood home was not a happy one. People expect too much of me. I am more temperamental than most people.)

Inadequate personality: 25 items (*e.g.*, I'd rather save my money than spend it on extra comforts. I have neither many friends nor any close friends. I don't have any special goal or ambition.)

Sexual psychopathy: 15 items (*e.g.*, There have been times in my life when I masturbated 4 or 5 times a day. I have had sexual intercourse very often. I have known a prostitute quite well.)

Nomadism: 15 items (*e.g.*, If I get tied down to any one place or job for too long, I feel nervous or trapped. I am not the kind of person who likes to travel. I like to change my surroundings often.)

"Other" psychopathic traits: 40 items (*e.g.*, I have no habits that are considered very bad. I find it hard to make new friends to take the place of old ones. I have never been up before a judge.)

The sub-scale items are randomly intermixed throughout the test. Although separate scores for sub-scales are readily obtainable, the present report will deal only with total test scores. Data for sub-scales cannot be analyzed meaningfully until separately diagnosed criterion groups are available for their validation.

Procedure

PEAQ scores were obtained from five different subject groups, two of which were considered psychopathic and three were normal controls. A total of 209 *Ss* were tested at different times over the past eight years. All testing was done in group situations and all *Ss* were assured that scores would not be available to personnel of their own institution or situation. All groups were told that we were obtaining samples for development of a psychological questionnaire. The five samples tested are briefly described below:

A. *Diagnosed reformatory sample.* Fifty-five inmates of the Ionia State Reformatory (Michigan) were tested during the Spring of 1949. These *Ss* had all been examined psychiatrically and those who had a primary diagnosis of psychopathic personality (with or without additional specification of type) were selected for the study. No other criterion of selection was applied. *Ss* ranged in age between 17 and 28, with a mean age of 21.00 $\pm$ 2.83 years. Their average education was 8.3 years and average IQ as determined by the prison psychologist was 96.38.

---

[2]*Personal Experience and Attitude Questionnaire.* Copyright 1956 by Mortimer H. Applezweig, Ph.D., Andrew S. Dibner, Ph.D., and Raymond L. Osborne, M. D. Available for research purposes from Dr. M. H. Applezweig, Connecticut College, New London, Conn.

B. *Diagnosed alcoholic sample.* Forty patients in the Blue Hills Clinic of the Connecticut Commission on Alcoholism (both in- and out- patients) were tested in early 1950. The PEAQ was administered as part of a regular testing battery to patients referred to the staff psychologist. A diagnosis of primary or secondary alcoholism was used as the criterion for inclusion of an *S*'s score in this sample. Hewitt[3], Buhler and Lefever[1] and Manson[6] all justify the inclusion of alcoholic *Ss* in a psychopathic classification. Psychometric data (MMPI) further supports this identification[6]. *Ss* in this group range in age between 22 and 58, with a mean of 38.15 ± 7.73 years. Educational level ranged from 3 to 17 years of schooling completed, with a mean of 9.5 ± 3.0 years.

A + B. The two samples described above, consisting of 95 *Ss*, constitute the psychopathic group.

C. *Trade School Sample.* In the spring of 1950, 32 students in the Wilbur Wright Vocational High School, a Detroit (Michigan) trade school, were tested as the first of three control samples. The age range of this sample was from 16 to 19 years, with a mean of 17.4 ± .82 years. Their education ranged from 9th to 11th grades, with a mean of 10.5 ± .05 years of schooling. This group was chosen because a portion of the reformatory group (sample A) also were brought up in Detroit.

D. *Navy enlisted men sample.* Twenty-four enlisted submarine candidates at the U. S. Navy Submarine Base in New London, Connecticut, were tested as part of a battery of experimental measures given in 1950 at the Medical Research Laboratory. The *Ss* in this group ranged in age from 17 to 23 years, with a mean of 19.17 ± 1.65 years. The mean level of schooling completed was 11.8 ± .76 years, with a range from 10 to 14 years.

E. *College sophomore sample.* A total of 58 college sophomore men were tested at Wesleyan University (Connecticut) in a classroom situation in psychology. The *Ss* were obtained in two successive years (1950, 1951), and the two samples combined when no differences were found between their scores. The College group ranged in age from 19 to 26 years, with a mean of 21.21 ± 1.88 years. All were in their 14th year of schooling.

C + D + E. The three samples, C, D & E constitute the 114 *Ss* of the control group. It should be noted that whereas a clear diagnosis of psychopathy or alcoholism was available in all cases of *Ss* included in groups A and B, respectively, no equivalent diagnoses of non-psychopathy were available for the control *Ss*. Their presence in the life situations in which they were found was taken as *prima facia* evidence of their representativeness of a typical adult male population, with no further investigation of freedom from psychopathy being made.

Results

Reliability. Table 1 presents Pearson coefficients for the two halves of the 140-item questionnaire, split odd-even within sub-tests. The right-hand column shows the same coefficients after application of the Spearman-Brown Prophecy Formula[2]. For all samples, and for samples combined, the measure is found to be highly consistent internally.

Table 1. Means, Sigmas, Ranges, and Split-Half* Reliability Coefficients for PEAQ Scores for Each Sample and for Psychopathic and Control Samples Combined.

Sample	N	Mean	σ	Range	r_{11}	r_{11}**
A (Reformatory)	55	60.35	20.47	14–106	.904	.950
B (Alcoholic)	40	58.43	18.79	16– 99	.885	.939
A + B (Psychopathic)	95	59.54	19.80	14–106	.897	.946
C (Trade School)	32	40.34	19.42	14–104	.899	.947
D (Navy Enl. Men)	24	38.25	20.30	15– 86	.917	.957
E (College)	58	32.97	12.96	7– 75	.844	.915
C + D + E (Control)	114	36.15	17.00	7–104	.886	.940
Total Sample	209	46.78	21.72	7–106	.922	.959

*Split odd-even within sub-scales.
**Spearman-Brown correction applied.

Comparison of psychopathic and control groups. Before comparing the psychopathic and non-psychopathic samples, Bartlett's test for homogeneity of variance[4] was applied, and the samples were found not to differ significantly in variance. Means, sigmas and ranges for total PEAQ scores for all groups are shown in Table 1. Table 2 shows results of an analysis of variance for the five samples treated separately

TABLE 2. ANALYSIS OF VARIANCE OF PEAQ SCORES FOR FIVE SAMPLES (N = 209)

Source	Sum of Sqs.	df	Variance Est.
Between groups	29686.02	4	7421.505
Within groups	68871.86	204	337.607
Total	98557.88	208	F = 21.98 P = .01

and Table 3 the *t*-test comparisons of each sample with each other sample. It can be seen from these tables that the groups differ significantly, with both psychopathic groups differing from all three control samples and not from each other. The differences between psychopathic and control samples are all significant at better than the .001 level of confidence.

TABLE 3. TESTS OF SIGNIFICANCE (*t*) OF DIFFERENCES BETWEEN GROUPS

Group	B	C	D	E
A (Reform.)	<1	4.42**	4.36**	8.47**
B (Alcohol.)		3.94**	3.97**	7.86**
C (Trade Sch.)			<1	2.13*
D (Navy enl. men)				1.39
E (College)				

*significant at the .05 level of confidence.
**significant at the .001 level of confidence.

The only other difference which was significant at all was the one between the Trade School and College samples, and this at the .05 level of confidence. It is not surprising to find such a difference between these two groups for they are close to opposite poles in the make up of a normal group with its wide range of behavior in conformity or non-conformity to society's standards. Examination of the differences between the groups indicated that the college group was higher principled and more controlled (less impulsive) than the trade school group, had fewer brushes with the law, was more considerate of other peoples feelings and less suspicious of others' motives. They also had fewer nomadic impulses and professed to fewer direct heterosexual experiences, although a greater percentage admitted to frequent autoerotic sexual practices. Despite these differences between the groups they are both considered as part of our normal sample.

Application of a cutting score. Following a procedure developed by Guilford and reported by Manson [6], we divided our total sample in half, by sub-samples, and determined a cutting score for psychopathy on half the population. The formula applied here was the following:

$$X_c = M_x + \frac{(.5 - p)\,\sigma^2_x}{M_p\,M_q\,(pq)}$$

Where X_c = cutting score, M_x = Mean of total sample used, M_p = Mean of the psychopathic group, M_q = mean of control group, and p and q the Ns of both samples, respectively. σ_x^2 = standard deviation of total sample.

The cutting score of 47 thus obtained was applied to the second half of the sample, resulting in correct identification of 73.2% of the control group and 70.8% of the psychopathic group. The overall correct identification was 72.1%. Since the number of "misses" was found to be higher than desired, recourse was made to an item analysis, and a first approximation of a test revision undertaken. In examining the relative percentages of psychopathic and control group responses to each item, it was found that 16 of the 140 items actually reversed the expectation that the psychopathic group would score higher. Several other items showed little or no differ-

ence between the two groups. The 16 reversing items were dropped from the total scores and a new cutting score derived. Applying the new cutting score improved control group identification to 80.4% and psychopathic group identification to 77.1%, with an over-all correct identification of 78.8%. It can certainly be expected that further elimination of non-discriminating items will improve the predictive efficiency of the measure. As with any use of clinical judgment as a criterion for the validation of a psychometric, a portion, at least, of the error remaining may be attributable to the lack of validity of the criterion. Since the control group may contain a few psychopathic individuals, the "false-positives" here may not all be false.

Discussion

Treating the 140-item psychopathy scale as a whole, we have found it to have high internal consistency, and to give evidence of validity for the discrimination of psychopathic from normal persons.

As indicated earlier, the PEAQ is composed of a series of items touching on varieties of asocial and anti-social behavior, from the more overt sexual and criminal acts to more subtle indicators of lack of impulse control, restlessness, lack of ethical standards, suspicion of authority, lack of goal persistence, lack of real personal ties, and so on. Whether these do indeed form a meaningful syndrome of "psychopathy" cannot be determined at this time. Use of an inventory such as the PEAQ and careful analysis of the data obtained with clinical groups should do much to help clarify the nature of any syndrome or syndromes involved.

Rubinstein and Lorr[9], for example, recently reported the successful use of a sample of 50 items taken from the PEAQ in the identification of patients who were to become early terminators in therapy. The items used by Rubinstein and Lorr were drawn largely from our Nomadism, Emotional Instability and Inadequacy scales. To a limited degree, their study offers evidence in support of the usefulness of the PEAQ in research on prediction problems where the so-called psychopathic traits of lack of impulse-control, hostility toward authority, restlessness and nomadism, and the level of goal-directed persistence toward culturally-approved goals seem crucial.

Summary and Conclusions

A 140-item true-false screening questionnaire for psychopathic behavior, the *Personal Experience and Attitude Questionnaire*, is described. Total score on the measure is highly reliable and discriminates at a highly significant level between diagnosed psychopathic and control samples. The use of a cutting score permits correct identification of 77% of the psychopathic group and 80% of a control sample. There is good promise of even better prediction with an empirically revised scale. The use of sub-scale scores for criminalism, emotional instability, inadequacy, sexual psychopathy and nomadism is considered possible only when validating information on these sub-scales becomes available.

References

1. Buhler, Charlotte and Lefever, D. W. A Rorschach study on the psychological characteristics of alcoholics. *Quart. J. Stud. Alcohol.*, 1947, *8*, 197-264.
2. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans, Green, 1947.
3. Hewitt, C. C. A personality study of alcohol addiction. *Quart. J. Stud. Alcohol.*, 1943, *4*, 368-386.
4. McNemar, Q. *Psychological statistics.* Second Edition. New York: Wiley, 1955.
5. Manson, M. P. A psychometric differentiation of alcoholics from nonalcoholics. *Quart. J. Stud. Alcohol.*, 1948, *9*, 175-206.
6. Manson, M. P. A psychometric analysis of psychopathic characteristics of alcoholics. *J. consult. Psychol.*, 1949, *13*, 111-118.
7. Maughs, S. B. Psychopathic personality: Review of the literature 1947-54. *Arch. crim. Psychodynamics*, 1955, *1*, 291-325.
8. Preu, P. W. The concept of psychopathic personality. In J. McV. Hunt (ed.). *Personality and the behavior disorders.* Vol. II. New York: Ronald Press, 1944.
9. Rubinstein, E. A. and Lorr, M. A. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, *12*, 345-349.

THE HILL-MONROE INVENTORY FOR PREDICTING ACCEPTABILITY FOR PSYCHOTHERAPY IN THE INSTITUTIONALIZED NARCOTIC ADDICT

JACK J. MONROE AND HARRIS E. HILL

U. S. Public Health Service Hospital, Lexington, Ky.

Problem

Several reports have appeared recently of attempts to select patients who would be continued in treatment by the therapist, and to predict the outcome of treatment[1, 2]. These studies were made on both outpatient and institutionalized groups and appear to have considerable promise for screening "drop-outs". The present paper reports an attempt to develop a self-administering inventory for selecting the individuals among a drug-addict population who are most acceptable for psychotherapy. Specifically, this paper (a) defines a criterion of acceptability for psychotherapy, (b) describes an inventory designed to predict this criterion, (c) presents the original cross-validation of combinations of items from the inventory, (d) indicates how the final acceptability scale was constructed, (e) includes normative data based on a large number of subjects and presents correlations between the scale and other variables, and (f) reports validity generalization tests of the scale on a different sample.

The Criterion

Since "acceptability for psychotherapy", as defined in a previous study[3], was employed as the criterion for the reliability and validity studies to be reported here, the original study will be described briefly. Each of three psychotherapists interviewed approximately one-third of 100 male narcotic addicts who comprised the subject group, and during three one-hour interviews " . . . , judged the acceptability of each patient for psychotherapy." These judgments constituted the criterion which was used for calculating the predictive power of a battery of psychological techniques. The psychotherapists found that the more acceptable patients showed a greater ability to verbalize feelings, spontaneously produced more problems for discussion, gave more evidence of original thinking, were more apt to ask for aid with their problems, and showed more evidence of previous ability to modify their patterns of living.

Description of the Inventory

The original pool of 180 items from which the inventory was derived was largely determined by hypotheses from other investigations of addict behavior[4, 5, 6] and from several years of clinical association with hospitalized addicts, on the supposition that the more "normal" individuals (in general, the less psychopathic) would be considered more acceptable for psychotherapy. Specifically, it was hypothesized that the more acceptable individuals would show (a) a greater degree of general motivation, *i.e.*, they would produce more positive responses to items having an "interest" or "satisfaction" content; (b) fewer positive responses to items which reflected the desirability of the drug condition; (c) less hostility toward authority and fewer antisocial and asocial attitudes; (d) less evidence of broken homes and extremely harsh or pampered rearing; (e) more sympathy and empathy; and (f) more interest in obtaining help in overcoming their problems. In order to minimize the semantic barrier, items of the inventory were written in language commonly used by addict patients[7].

Cross validation on original sample. The inventory was administered in small groups to 100 institutionalized addict patients who were subsequently judged acceptable or unacceptable for therapy. Upon termination of the study the group was divided into random halves to form two subgroups. An item count for each

sub-group was made to isolate the items which were significantly related to the criterion. Using items which produced t values for the difference between proportions that exceeded 1.5, empirical scoring keys were constructed to reflect the direction of response (true or false) of the acceptable subjects. The key developed for one subgroup was used to predict acceptability for psychotherapy in the other subgroup, and vice versa.

This procedure produced agreement between the inventory results and the judgment of the therapists in 74 percent of all cases. If the inventory had been used to refer to therapy those patients scoring in the upper quartile, 78 percent of such referrals would have proved acceptable for psychotherapy as indicated by the criterion.

Construction of acceptability scales. Seventy-eight of the 180 items of the inventory yielded the same direction of response in the two sub-samples and produced tetrachoric correlation coefficients with the criterion which were statistically significant at the .05 level of confidence. Two senior psychiatrists of the hospital staff, neither of whom had participated in the initial study, scored the 180 items to reflect

TABLE 1. ITEMS OF ACCEPTABILITY SCALE C, SHOWING DIRECTION OF RESPONSE OF ACCEPTABLE SUBJECTS

I usually have a good appetite. (T)

Many well-adjusted people become addicted. (F)

The only thing wrong with addiction is that there is a law against it. (F)

I feel most like quitting drugs for good when kicking a habit. (F)

I have never been in love with anyone. (F)

Marriage has only (or would only have) interfered with my activities. (F)

There should be more community clinics to treat addicts after they leave the hospital. (T)

There is nothing wrong with me (except that I use drugs). (F)

Many of my dreams are about sex. (T)

As long as I can have drugs I am not interested in help to overcome my habit. (F)

I think most people tell lies to get ahead. (F)

I am a good mixer. (T)

I wish I did not have to associate with addicts. (T)

I had strong affection for my mother. (T)

There is nothing I could not do if I had drugs all the time. (F)

I know I am different from the normal man but I don't know what the difference is. (T)

I finished high school. (T)

All addicts are pretty much alike. (F)

I am in favor of making drug use legal for real addicts. (F)

The main reason for coming to this hospital was that my habit was getting too big and too expensive. (F)

I had psychiatric help on the outside. (T)

I am likely not to speak to people until they speak to me. (F)

I like to sit around and enjoy a shot. (F)

I will never be satisfied with a routine job. (F)

I liked school. (T)

I think a psychiatrist could help me. (T)

Most people are honest because they are afraid of being caught. (F)

Most people make friends because friends are likely to be useful to them. (F)

Our society should be changed so that an addict can get what is due him. (F)

I wanted to learn a profession but something always came up to stop me. (F)

I have had some education beyond high school. (T)

Drug addiction can be inherited. (F)

Any addict with will power should be able to give up drugs on his own. (F)

I wish I were not so shy. (T)

My parents were living together when I left home (or, are still living together, if you still live with them). (T)

I played hooky from school quite often as a youngster. (F)

The problem of drug addicts would be solved if doctors could develop a drug which relieves pain, but is not habit-forming. (F)

If a doctor were called to see me when I was seriously ill (and did not have a habit), I would tell him that I had been addicted to drugs and for him to be careful with what he gave me. (T)

A drug addict is more respectable than an alcoholic. (F)

An addict never does abnormal sex acts when on junk. (F)

Sometimes I take myself too seriously. (T)

I must admit, I wish I were a "Square John". (T)

I rarely have intercourse when I am not taking drugs. (F)

There are very few things that interest me. (F)

If I have been off drugs for awhile, I feel a little sick just before taking the first shot. (T)

the direction of response which they judged would characterize acceptable therapeutic candidates. It was found that of the 78 items that empirically differentiated acceptable and non-acceptable subjects in the original cross-validation, 46 were agreed upon by the psychiatrists. These items (scale C), which met both empirical and rational requirements, are presented in Table 1 together with the direction of response of the acceptable patients. For calculating indices of reliability for this scale, the 46 items were used to construct two 23-item scales (Scales A and B) by assigning items to each that were paired for both content and validity coefficient.

Normative data and correlations with other variables. The inventory was administered to 400 institutionalized narcotic addicts who remained in the hospital a sufficient period to become screening problems. The distribution of raw scores on Scale C is presented in Figure 1. As can be seen, these scores approximate a normal

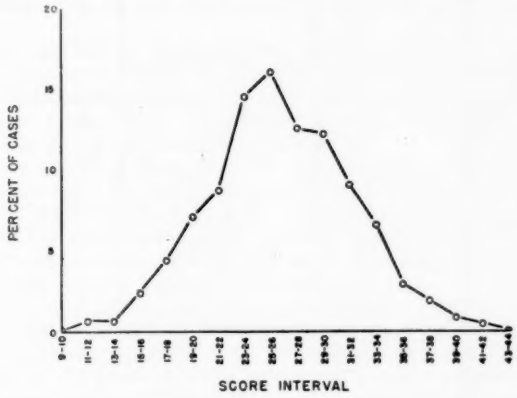

FIGURE 1. FREQUENCY DISTRIBUTION OF RAW SCORES ON ACCEPTABILITY SCALE C OBTAINED BY 400 INSTITUTIONALIZED NARCOTIC ADDICTS

distribution, with the fiftieth percentile falling at 26, the eighty-fourth at about 31, and the sixteenth at 21. The scaling methods described by Johnson[8] were then used to transform the raw data to T-scores giving a normal distribution with a mean of 50 and a standard deviation of 10. Raw scores and equivalent T-scores are presented in Table 3.

A product-moment coefficient of reliability of .67 (N = 400) was obtained for scale C by correlating Scale A with Scale B and converting to full length by using the Spearman-Brown prophecy formula. Reliability of the full-length scale is somewhat greater in sub-samples of more highly motivated subjects. In 101 patients, who had requested psychotherapy, a product-moment coefficient of .75 was computed. Furthermore, reliability of the 46-item scale is not affected by removing "buffer" questions from the inventory. A brief questionnaire comprised of the 46 keyed items, plus 14 others designed to measure test-taking attitude, was administered to a second sample of 101 patients who had requested psychotherapy, and a product-moment correlation coefficient of .73 was obtained.

All patients in this study were male. The applicability of the scale to female addicts has not been tested. The differences between the means of White and Negro subjects were statistically negligible. However, there was a statistically significant difference between the standard deviations of these racial groups, the negroes showing less variability. A tetrachoric coefficient of correlation between chronological age and acceptability test score in 200 patients was +.30; this was statistically significant beyond the .01 level of confidence. In the same sample, however, the co-

TABLE 3. ACCEPTABILITY TEST SCORES OBTAINED BY THREE GROUPS OF INSTITUTIONALIZED NARCOTIC ADDICTS. (T-SCORES BASED ON RESPONSES OF N = 400.)

T-Scores	Raw Scores	In Individual Therapy (3 or more interviews)	Group Therapy (3 or more interviews)	Not Selected for Therapy or "dropped out"
73–74	38–39	2		
71–72	37	4	1	
69–70				
67–68	35–36	4	2	
65–66	34	2	2	1
63–64				
61–62	32–33	8	8	2
59–60	31	1	3	1
57–58	30	3	1	2
55–56	29	2	4	4
53–54		2	3	4
51–52	27	2	4	2
49–50	26	2	2	4
47–48	25	1		5
45–46	24	3	3	5
43–44	23	2	1	5
41–42	22	1	3	3
39–40	20–21	2	2	6
37–38	19		1	2
35–36	18		2	
33–34	17	2		
31–32	16		1	2
29–30				
27–28	15			1
25–26	13–14			
23–24	12			
21–22				
19–20	11			1
		N = 43 M = 56.72 SD = 10.92	N = 43 M = 52.88 SD = 9.83	N = 50 M = 46.68 SD = 8.86

efficient of correlation between age and selection-rejection for therapy by means of psychiatric staffing was nonsignificant (–.09).

A statistically significant relationship was found between the acceptability scale and certain test variables of the MMPI [9] and the Cattell Sixteen Personality Factors Test [10] as shown in Table 2. The significant tetrachoric correlations found between the scale and the more "healthy" variables, such as Responsibility, Mf, and Cattell's Factors A and N, tend to support the writers' hypothesis that the less psychopathic and more normal subjects would score higher on acceptability for psychotherapy. Even the significant correlations found between the scale and Neuroticism add support to this position if it is assumed that, in general, the neurotic and anxious individuals in the addict population profit more from therapy than the individuals labelled psychopathic character disorders.

Validity generalization. Between October 1955 and April 1956 the psychiatric staff of the hospital selected 94 patients for psychotherapy independently of scores on the acceptibility scales. Eight of these patients were dropped from therapy within one or two interviews, 43 of the remaining 86 patients were assigned to individual psychotherapy, and 43 were assigned to group psychotherapy. These subjects constituted the experimental groups in the validity generalization study. The inventory

TABLE 2. CORRELATIONS OF THE ACCEPTABILITY SCALE WITH OTHER TESTS

MMPI Scales (N = 93) *Test*	r_t	Cattell 16 P. F. Test Factors (N = 129) *Factor*	r_t
Responsibility	+.50**	A. Warm, sociable	+.30*
Mf	+.42**	N. Sophisticated, polished	+.30*
Neuroticism	+.40**	I. Sensitive, effeminate	+.19
Dominance	+.30	B. Bright, intelligent	+.16
Anxiety Index	+.29	Q^1 Experimenting, critical	+.14
Psychopathic deviate	+.15	G. Conscientious, persistent	+.10
Hypomania	−.20	M. Eccentric, unconcerned	−.10
Hypochrondrism	−.29	Q^4 Tense, excitable	−.10

*Significant at the .05 level of confidence
**Significant at the .01 level of confidence

was administered routinely to these subjects, and was administered concurrently to 42 other patients who were admitted to the hospital during the same period of time and who received psychiatric staffing but were not selected for psychotherapy. These 42 patients, plus the 8 who were dropped from therapy within one or two interviews, provided the control group for therapeutic selectees.

The predictive utility of Scale C can be calculated from Table 3. Although the mean T-score for the group which was chosen for individual therapy was slightly higher than that of the group chosen for group therapy, the difference was not statistically significant. The mean scores for the therapy groups, however, were significantly higher than that of the controls, the differences between means being statistically significant at the .001 and .01 percent levels respectively.

If it is assumed that the ratio of acceptable to unacceptable therapeutic candidates is 43 to 50 (.46),[1] the data in Table 3 provide an estimate of the relative efficiency of the scale in selecting patients for individual and group therapy. A cutting point of 56 "picks up" 58 percent of the individual therapy assignees, and 40 percent of the group therapy assignees, while picking up only 12 percent of the controls. Seventy-seven percent of a sample of patients selected in this way would be acceptable for either individual or group therapy in contrast to the acceptability base-rate of .46 in the addict population. If a cutting point of 60 were used, 85 percent of the test-selected sample would be acceptable for psychotherapy.

The present inventory was designed specifically for use with narcotic addicts. However, a preliminary investigation which used parallel forms of this inventory with addicts and alcoholics has indicated that it may be applicable to the study of the alcoholic. Differences were found between the two groups in responses to items designed to measure sensitivity to pain, kind and degree of hostility and aggression, as well as between items concerning preferred drug effects.

SUMMARY

A questionnaire of 46 items was constructed out of a pool of 180 items for selecting individuals among a drug-addict population who were most acceptable for psychotherapy. The criterion of acceptability was the judgment of psychiatrists made during three one-hour interviews with each patient. The first study utilized 100 institutionalized narcotic addicts after withdrawal treatment had been completed.

An item count was made to find the items which were significantly related to the criterion. Cross validation demonstrated that combinations of items from the

[1]In a recent study[3] psychiatrists estimated that 46 per cent of the addict patients that were interviewed were acceptable for psychotherapy. Thus, each therapy group (Table 4) should be combined with the control separately to calculate per cent efficiency above a particular cutting point. If therapy groups are combined, as was done above, the control group should be doubled, since it is assumed that 50 unacceptable candidates were rejected in the process of selecting each of the therapy groups.

questionnaire were 74 percent efficient in predicting psychiatric judgments of acceptability for psychotherapy. For purposes of further refinement, 78 items that differentiated between the more acceptable and the less acceptable patients in the initial study were reduced to 46 by selecting only those empirically valid items which two senior psychiatrists, neither of whom had participated in the original cross-validation, independently judged to reflect the responses of acceptable therapeutic candidates.

Normative data for the inventory on 400 subjects were obtained. Validity generalization tests of the predictive efficiency of the 46 item scale were then made on a group of 136 subjects, some of whom had been independently chosen to receive psychotherapy, and some of whom had not been so chosen. When the scale was used with this sample to label those patients scoring more than one standard deviation above the population mean as "acceptable for therapy" most were those for whom psychotherapy had been prescribed.

References

1. Barron, F., An ego strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, *17*, 327-333.
2. Wirt, R. D., Further validation of the ego strength scale. *J. consult. Psychol.*, 1955, *19*, 444.
3. McLean, A., Monroe, J., Yolles, S., Hill, H., and Storrow, H., Acceptability for psychotherapy in institutionalized narcotic adicts. *Arch. Neurol. Psychiat.*, 1955, *74*, 356-362.
4. Hill, H. E., Belleville, R. E., Haertzen, C. A., and Glaser, R., Personality differences of narcotic addicts and alcoholics as indicated by the MMPI. *Quart. J. Stud. Alcohol.* In press.
5. Kolb, L., Types and characteristics of drug addicts. *Mental Hygiene*, 1925, *9*, 300-313.
6. Felix, R., An appraisal of the personality of the addict. *Amer. J. Psychiat.*, 1944, *100*, 462-467.
7. Maurer, D. W., and Vogel, V. H., *Narcotics and Narcotic Addiction.* Springfield, Ill.: Charles C. Thomas, 1954, pp. 303.
8. Johnson, P. O., *Statistical Methods in Research.* New York: Prentice-Hall, 1949.
9. Hathaway, S. R. and McKinley, J. C. *Minnesota Multiphasic Personality Inventory Manual.* New York: The Psychological Corp., 1951.
10. Cattell, R. B. *The Sixteen Personality Factors Test Manual.* Champaign, Ill.: Cattell, Univ. of Ill., 1950.

A NEW SORTING TECHNIQUE FOR DIAGNOSING BRAIN DAMAGE[1]

JOHN ALTROCCHI AND B. G. ROSENBERG
University of California[2] *Bowling Green State University*

Problem

Since the pioneering work of Goldstein [8], evaluation of abstract behavior by means of sorting ability has been of central importance in the psychodiagnosis of organic brain damage [9]. There is, however, a scarcity of quick, easy, quantitative sorting tests available [14, 19]. This report concerns an investigation of a new sorting technique which may help meet this need.

The technique was developed by Yacorzynski [17, 18]. There are 16 blocks which the subject is asked to sort into four groups of four blocks each.[3] The unique feature

---

[1]Statistical analysis was facilitated by a grant from the Scholarly Advancement Committee, Bowling Green State University.

Most of this investigation was carried out at the V. A. Hospital, Palo Alto, Calif., and the V. A. Hospital, Oakland, Calif. We gratefully acknowledge the considerable advisory help of Drs. Albert Kostlan, Paul McReynolds, William Grove, and Richard Sears.

[2]Now at Duke University.

[3]Dittoed copies of instructions and scoring criteria developed by the authors can be obtained by writing the senior author, Department of Psychiatry, Duke University.

of the test is that the blocks can be sorted eight different ways: (1) shape of block, (2) shape of figure (inscribed on the top of the block), (3) color of block, (4) color of figure, (5) height, (6) size or volume, (7) area of the top of the block, and (8) area of the figure. The subject is asked to keep sorting the blocks in new ways until he gives up, spends ten minutes without achieving a sorting, or repeats a category five times. Yacorzynski suggested that non-organics over ten years of age can achieve at least four sortings and that many organics cannot achieve four or even three.

The authors undertook extensive pretesting of the technique. It was found to be interesting to patients and administrable in an average of twenty minutes. A detailed but objective scoring system was developed to give partial credit for partial solutions and to take account of complexity of verbalization. Scores range from zero to eight in one-third-unit steps (*i.e.*, three-plus, four-minus, four, four-plus, etc.). Administration of the technique to 18 hospitalized organic patients, 14 hospitalized psychiatric patients, 13 aphasic outpatients, and 21 normals suggested that the technique was reliably able to separate organics from others. Therefore, a more rigorous investigation was undertaken.

First Validity Study

The first study was designed to discover if the test could differentiate organics from schizophrenics[11]. This study was conducted in a large V. A. neuropsychiatric hospital. In view of evidence from recent investigations[1, 3, 4, 5, 13] that "process" or long-term, chronic schizophrenics may have organic damage, as differentiated from "reactive" or acute schizophrenics, three groups of subjects were needed. A group of reactive schizophrenics was unfortunately not available. The organic group was composed of 19 hospitalized male patients with a primary diagnosis of syphilis, tertiary, meningo-encephalitis (paretics). The ward physician individually selected 19 schizophrenic patients (all presumably of the process type) to match the organic patients in age, length of illness, and severity of malfunctioning.

Since it seemed necessary to discover the effect of intelligence on the Yacorzynski Block Technique (YBT), three Wechsler-Bellevue (WB) subtests were selected which seemed to give broadest coverage, maximum reliability[10], and to be minimally affected by brain damage[12, 14]: Information, Comprehension, and Picture Completion. A part-whole correlation was run between the total weighted scores on the three subtests (hereafter referred to as WS) and WB total scores for a sample of 74 mixed psychiatric patients.[4] The resulting correlation of .89, albeit spurious, was sufficient to suggest that these three subtests estimate WB IQ for this kind of population with satisfactory accuracy.

The subjects were administered the three WB subtests and the YBT "blind", so that the administrator of the YBT did not know whether the subject was organic or schizophrenic.

Results. The results of the matching and testing are shown in Table 1. Age was successfully matched. Length of illness was significantly different for the two groups, but the difference between 20 and 27 years of illness suggests that this was not a differentiating factor. Though the schizophrenic group did slightly better on the YBT, the difference is not significant, most patients in both groups performing

Table 1. Results of the First Study

Group	Mean Age	Mean WS†	Length of Illness (years)	Concepts X	Concepts S.D.
Organic (N = 19)	59.1	18.3	19.68	1.789	1.472
Schizophrenic (N = 19)	60.0	21.4	27.26*	2.579	1.350

*Significant at the .01 level of confidence. †Total weighted scores of three subtests.

[4]We are grateful to Dr. W. T. Smelser for making these data available to us.

within the suggested organic range. Thus, if these elderly, deteriorated schizophrenics were indeed non-organic, the technique failed in its purpose. If, however, as seems more likely[1, 3, 4, 5, 13] they should be viewed as organic, the technique performed adequately so far. It needed to be demonstrated that non-organics score higher on the YBT.

Second Validity Study

The Chief Neurologist of a large V. A. General Medical and Surgical Hospital[5] selected from the neurology ward forty patients who were either demonstrably organic (central nervous system pathology rostral to the foramen magnum) or non-organic (no such pathology).[6] Each patient had had a complete neurological work-up, including neurological examination and routine laboratory procedures and, when necessary, EEG, lumbar punctures, and skull X-ray. These patients were seen "blind" by the authors and given the three WB subtests and the YBT. The authors did not know the proportion of organic versus non-organic patients until the completion of the study. This method is similar to that of Fisher, Gonda, and Little[6], the research situation being as representative of the clinical situation as possible. The design follows the suggestion of Piotrowski[16] that the best control group for a group of organics is a group of patients who were suspected of organic damage but later found to be non-organic.

Results. The criterion yielded a group of 28 organics and 12 non-organics. Significant differences were not found in *age* (organic: $\bar{X}$ = 40.39 years, S.D. = 13.50; non-organic: $\bar{X}$ = 40.16 years, S.D. = 11.61) or *intelligence estimate* (organic: $\bar{X}_{WS}$ = 31.04, S.D. = 5.98; non-organic: $\bar{X}_{WS}$ = 32.50, S.D. = 5.73). Table 2 presents the agreement of the YBT and the criterion diagnosis. In view of lack of empirical norms, three sets of predictions were made from the data before the criterion was known: (a) *Difference analysis*, where patients were diagnosed organic if their YBT score was lower than expected from their intelligence estimate; (b) *Difference plus qualitative*, using the difference method plus any qualitative signs that had been noted during the administration of the YBT; (c) *Cut-off four-minus*, a simple mechanical method ignoring intelligence estimate and diagnosing as organic anyone scoring below four-minus. After the criterion was known, a fourth method was tried:

Table 2. The Relationship Between Four Interpretive Systems for the YBT and the Criterion Diagnosis for the Second Study

		Criterion		
		NO	O	Sum
Difference Analysis	O	1	15	16
	NO	11	13	24
		12	28	40

ϕ = .408
P = .01

		Criterion		
		NO	O	Sum
Difference + Qualitative	O	1	14	15
	NO	11	14	25
		12	28	40

ϕ = .394
P = < .02

		Criterion		
		NO	O	Sum
Cut-off Four-Minus	O	2	15	17
	NO	10	13	23
		12	28	40

ϕ = .342
P = < .05 > .02

		Criterion		
		NO	O	Sum
Cut-off Four	O	3	20	23
	NO	9	8	17
		12	28	40

ϕ = .430
P = < .01

[5] We are extremely grateful to Dr. J. K. Smith of Oakland V. A. Hospital for his careful handling of this crucial role in the study.

[6] We realize that this is an arbitrary dividing line, but current neurological knowledge and research precedent seem to make it a reasonable one

(d) *Cut-off four*, whereby any score less than four was considered organic. As can be seen from Table 2, all four systems correlated significantly with the criterion. Interpreting Table 2, all four systems agree highly with the criterion when they identify a patient as organic (94%, 93%, 88%, and 87%). When the YBT fails to identify a patient as organic the results are inconclusive (46%, 44%, 43%, and 53%).

The data were also analyzed for information on the influence of age and general intelligence on YBT score. For the second study the correlation between YBT score and age was −.199 (N = 40, P = >.10), so that age does not appear to be of critical importance. The correlation between YBT score and WS estimate of intelligence was .69 for the first study, where the range of both scores was restricted, and .56 for the second study where the range was more representative. It is worthy of note that despite the significant (P = <.01) relationship, WS alone failed to distinguish between the organic and non-organic groups in the second study, and diagnosis on the basis of difference between YBT and WS (difference analysis) was not significantly more accurate than diagnosis from the YBT alone.

One further type of analysis was undertaken. Table 3 shows that non-organics take less time to achieve their first and second concepts. It may therefore be possible, as more normative data is collected, to increase the diagnostic accuracy of the test by weighting the scoring for time used.

TABLE 3. TIME (IN SECONDS) REQUIRED TO ACHIEVE CONCEPTS IN THE SECOND STUDY

YBT Concepts	Organic	Non-organic	*t*	P
1	X = 203.33 N = 28	X = 101.58 N = 12	2.055	<.05
2	X = 418.41 N = 27	X = 258.16 N = 12	4.981	<.01
3	X = 445.80 N = 20	X = 396.75 N = 12	.7174	
4	X = 540.25 N = 12	X = 480.60 N = 10	.2450	
5	X = 608.00 N = 5	X = 628.75 N = 4		

As a step toward the development of norms, the distribution of YBT scores for all subjects in both studies is presented in Table 4.

TABLE 4. FREQUENCY OF SCORES ACHIEVED BY ALL SUBJECTS ON THE YBT

Groups	YBT Score														
	0+	1−	1	1+	2−	2	2+	3−	3	3+	4−	4	4+	5−	5	5+
Paretic N = 19	3	3	1	2	3	2	2				2				1	
"Process" Schizophrenic N = 19	1	1	2		4	1	1	4			2	1		1	1	
Organic N = 28	1		1		4	2		5	2		5	3		2	3	
Non-organic N = 12								1	1		1	5		1	3	

DISCUSSION

The YBT is a relatively short, simple, interesting, and unthreatening method for evaluating sorting ability. The range of solutions allows for "success" for low-level subjects and challenge for even the brighest subjects. The scoring is objective,

quantitative, and requires little training. The technique also allows for observation of many qualitative cues, such as repetition of sortings, for evaluating a patient's handling of a concept formation task. The YBT, even without control for intelligence, discriminated many organics from other patients while misclassifying very few non-organics as organic. A test with this characteristic may be included in a battery of tests for organicity without the danger of many false diagnoses. The use of the conservative cut-off of four-minus is recommended for clinical situations. This tentative demonstration of the YBT's promise, however, calls for extensive cross-validation before it can be established as a valuable diagnostic tool.

Summary

The Yacorzynski Block Technique allows a subject to use eight categories of shape, color, size, and area in sorting 16 blocks into four groups of four. The test is objectively scored by the number of sortings achieved. Tentative norms and scoring criteria were developed in pretesting. The YBT and three WB subtests were administered to four groups of hospitalized male patients. The technique did not discriminate between 19 deteriorated "process" schizophrenics (sometimes considered organic) and 19 deteriorated paretics matched for age and severity of malfunctioning. It significantly discriminated, in a manner useful for clinical diagnosis, 28 organic from 12 non-organic patients on a neurology ward. Time to achieve the first two sortings was significantly different for the last two groups. The test thus tentatively appears useful in the detection of organicity.

References

1. Bellak, L. *Dementia Praecox.* New York: Grune & Stratton, 1948.
2. Bender, L. Psychological principles of the Visual Motor Gestalt Test. *Trans. N. Y. Acad. Sci.*, 1949, 11, 164-170.
3. Brackbill, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, 53, 210-226.
4. Brackbill, G. A. and Fine, H. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, 52, 310-313.
5. Bychowski, G. Physiology of schizophrenic thinking. *J. nerv. ment. Dis.*, 1943, 98, 368-386.
6. Fisher, J., Gonda, T. A. and Little, K. B. The Rorschach and central nervous system pathology: a cross-validation study. *Amer. J. Psychiat.*, 1955, 111, 487-492.
7. Goldstein, K. and Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2, (Whole No. 239).
8. Grassi, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology.* Springfield, Ill.: Charles C. Thomas, 1953.
9. Halstead, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and nonequivalent stimuli. *Amer. J. Psychiat.*, 1940, 96, 1263-1294.
10. Hamister, R. C. The test-retest reliability of the Wechsler-Bellevue Intelligence Test, (Form I) for a neuropsychiatric population. *J. consult. Psychol.*, 1949, 13, 39-43.
11. Hanfmann, E. and Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.
12. Herring, F. H. An evaluation of published short forms of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1952, 16, 119-123.
13. Kantor, R. E., Wallner, J. M. and Winder, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, 17, 157-162.
14. Klebanoff, S. G., Singer, J. L. and Wilensky, H. Psychological consequences of brain lesions and ablations. *Psychol. Bull.*, 1954, 51, 1-41.
15. Piotrowski, Z. A. The Rorschach inkblot method in organic disturbance of the central nervous system. *J. nerv. ment. Dis.*, 1937, 86, 525-537.
16. Piotrowski, Z. A. Positive and negative Rorschach organic reactions. *Rorschach res. Exch.*, 1940, 4, 147.
17. Yacorzynski, G. K. Concept formation as a function of personality structure. (Paper read at Amer. Psychol. Assoc. meetings, Sept., 1949. Abstr. in *Am. Psychol.*, 1950, 5, 322.)
18. Yacorzynski, G. K. Personal Communication, 1954.
19. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, 51, 359-379.

AN EVALUATION SCALE FOR SHOCK THERAPY

MARVIN J. FELDMAN

University of Buffalo

The present study describes the empirical derivation of an objective test designed specifically to evaluate the results of shock treatment. It may, however, be useful in assessing any somatic therapy with psychotic patients. Further it may enhance our knowledge of the verbal changes which accompany successful treatment.

Derivation of the Evaluation Scale

The study was part of a larger project to construct empirically both prognosis and evaluation scales for shock treatment. The Minnesota Multiphasic Personality Inventory (MMPI), given before and after treatment, furnished the initial pool of items for constructing the two scales. The prognostic scale (Ps) [2] has been described in detail elsewhere. The evaluation scale complements the work on prognosis. The same groups were used in constructing both scales except that some patients were not tested after shock so had to be eliminated as subjects for the evaluation scale. Certain procedures for selecting cases are common to both scales. Since the article on prognosis describes the selection and personal characteristic of the groups in detail, only the general outline of these matters will be repeated in the present study.

The general plan for the evaluation scale was to select criterion groups of *recovered* and *unimproved* patients. An item analysis was performed on the pre- and post-shock MMPI records of these groups. The item analysis was somewhat unusual in requiring the following three steps:

1. An initial pool of items was selected which differed significantly between the pre-and post-shock records of the *recovered* group. From this pool, items were eliminated for either of the following two reasons.
2. There was also a significant change in the same direction in the pre-and post-shock *unimproved* group.
3. There were no significant differences between the item counts in the post-shock *recovered* group and the post-shock *unimproved group.*

This procedure should produce items reflecting changes in verbal behavior accompanying improvement.

The *recovered* and *unimproved* criterion groups had been classified on the basis of both psychiatric ratings at the conclusion of therapy, just prior to discharge from the hospital, and one year follow-up ratings. In selecting cases, a restriction was imposed to the effect that the pre-shock MMPI record had to deviate sufficiently from normality. The groups were fairly typical of hospital patients used in studies of this sort. They were slightly above average in intelligence; the average age was in the middle thirties; there was a slight preponderance of men over women, and the course of treatment was similar in both groups. The greatest discrepancy was in terms of diagnosis; the *recovered* group contained a preponderance of affective disorders, the *unimproved* group a preponderance of schizophrenics.[1]

The two criterion groups each contained thirty cases. From the pre- and post-shock MMPI records, item counts were tabulated separately for each group. After this tabulation, the first step was to cull items which changed by 23.3% (approximately the five percent level) in the pre- to the post-shock records of the *recovered* group. Of the total 550 items, 131 changed in this manner. Eight of the 131 items were eliminated because they also changed significantly from pre- to post-shock in

[1]The reader is referred to the monograph on prognosis (2) for an extended description of the selection and composition of the criterion groups.

the *unimproved* group. Another 55 items were eliminated because they did *not* differ significantly from the item counts in the post-shock records of the *unimproved* group. The five percent level was used for each of the latter two steps. The remaining 68 items comprised the first edition of the evaluation scale. The scale was scored so that a high score accompanied improvement.

Validation of the Evaluation Scale

The first step in establishing validity was to apply the scale back to the criterion groups. Two conditions must hold if the test is to be discriminating: (a) A significant difference between pre- and post-shock scores of the *recovered* group, and (b) a significant difference between the post-shock scores of the *recovered* and *unimproved* groups. Table 1 shows that both requirements are fulfilled satisfactorily in the criterion groups, for the evaluation scale separates the groups with a high degree of success. Table 1 also indicates that the pre-shock scores of the *recovered* and *unimproved* groups were quite similar and scores of the *unimproved* group do not change on the post-shock test. These latter findings would seem to reflect desirable properties of an evaluation scale.

Table 1. Application of the Evaluation Scale to the Criterion Groups

Criterion Groups	Mean	SD	diff.	CR	P
30 post-shock recovered	55.2	9.0	24.6	7.29	.0001
30 post-shock unimproved	30.6	14.7			
30 pre-shock recovered	29.8	12.4	25.4	8.70	.0001
30 post-shock recovered	55.2	9.0			
30 pre-shock unimproved	24.0	11.0	6.6	1.94	.05
30 post-shock unimproved	30.6	14.7			
30 pre-shock recovered	29.8	12.4	5.8	1.87	.06
30 pre-shock unimproved	24.0	11.0			

The next step was to cross validate the scale on new cases. Of the original one hundred cross validation cases used in the prognosis study, only sixty-three had post-shock MMPI's. These cases were similar to the criterion cases except that follow-up periods as short as six months were used in conjunction with the psychiatric ratings at the conclusion of treatment to assign patients their recovery status. The cross validation cases also included an intermediate group of cases rated *improved*. The comparisons of the results of the post-shock scores of the cross validation cases are shown in Table 2.

The evaluation scale has some degree of success in separating the *recovered* group from each of other two groups, but this unbiased estimate of validity produces considerably lower critical ratios compared with results from the criterion groups. Scores of the *improved* group are higher, but not significantly so, than the *unimproved* group. This latter finding is consistent with results obtained from the prognosis scale. The *improved* group may be closer to the *unimproved* than the *recovered* group, clinically as well as psychometrically. That is, the *improved* group may be labeled such because they make a satisfactory social adjustment. Nevertheless they may still retain a deep layer of psychopathology so that upon closer analysis they may not differ too much from the *unimproved* group. An evaluation scale would be more useful, however, to the degree that it can make finer discriminations among groups. The value of the present scale is impaired by its inability to separate scores of *improved* and *unimproved* groups.

TABLE 2. APPLICATION OF EVALUATION TO THE POST-SHOCK MMPI RECORDS OF THE CROSS VALIDATION GROUPS

Test Groups	Mean	SD	diff.	CR	P
26 recovered	53.9	6.9	19.2	4.47	.0001
11 unimproved	34.7	12.6			
26 recovered	53.9	6.9	13.2	4.63	.0001
26 improved	40.7	13.4			
26 improved	40.7	13.4	6.0	1.28	.20
11 unimproved	34.7	12.6			

An inspection of the data reveals that the *recovered* group are separated most accurately with little variability. The *improved* cases vary widely on the scale which may reflect the clinical heterogeneity of this group. A sizable minority of *unimproved* cases have favorable scores thus impairing the value of the scale. Perhaps these patients register transient gains in mental health which result in high scores.

FINAL ITEM ANALYSIS

Previous work [4] suggests that gains in validity are potentially marked as criterion groups are increased from 20 to 40 cases. Hence a final form of the scale was constructed with expanded criterion groups. Ten former cross validation cases were used to increase the *recovered* and *unimproved* groups to 40 cases in each group. The item analysis produced 62 items of which 44 overlapped with items on the preliminary scale. The final scale was labeled "EVS" to signify its purpose, "evaluation of shock therapy."[2] All items on both forms are answered in the "plus" direction by the *unimproved* groups and of course in the opposite direction by the *recovered* group. A "plus" answer is by definition the statistical minority response of the original normative sample for the MMPI. The 62 items are listed in Table 3 according to their numbers as given in the back of the revised manual of directions, 1943. [5]

TABLE 3. THE 62 MMPI ITEMS COMPRISING THE "EVS" SCALE. T OR F INDICATES THE DIRECTION OF RESPONSE BY PATIENTS WITH UNFAVORABLE PSYCHIATRIC STATUS AFTER THERAPY

A-8 (F)	F-10 (T)	F-55 (T)	I-23 (T)
A-12 (T)	F-11 (F)	G-1 (T)	I-27 (T)
A-26 (T)	F-15 (F)	G-2 (T)	I-30 (T)
A-40 (T)	F-17 (T)	G-4 (T)	I-31 (T)
A-47 (T)	F-35 (T)	G-6 (T)	I-32 (T)
B-27 (F)	F-36 (T)	G-7 (T)	I-34 (T)
B-28 (T)	F-37 (F)	G-12 (F)	I-35 (T)
B-30 (F)	F-39 (F)	G-36 (T)	I-37 (T)
C-17 (F)	F-41 (T)	G-51 (T)	I-39 (T)
E-27 (T)	F-43 (T)	H-1 (T)	I-54 (T)
E-29 (F)	F-45 (T)	H-13 (T)	J-4 (T)
E-36 (F)	F-46 (T)	H-28 (T)	J-6 (T)
E-43 (T)	F-48 (F)	H-31 (F)	J-13 (T)
E-44 (T)	F-49 (T)	I-10 (T)	J-33 (F)
E-50 (T)	F-50 (T)	I-12 (T)	
F-6 (F)	F-54 (T)	I-13 (T)	

EVS discriminates slightly better than the preliminary scale when applied back to the original criterion cases. No real cross validation was possible with the few remaining cases. It is reasonable to assume, however, that the validity of EVS is at least as great as the first form. Nevertheless, the value of EVS would seem to be greater for the light it may throw on the process of improvement rather than on its possible use as a clinical instrument.

[2]Copies of the scale may be obtained by request from the author.

Item Overlap with Other Scales

The items on EVS overlap considerably with other MMPI scales, for as shown in Table 4 there is: (a) Considerable overlap with items on the neurotic scales especially D, Hy, and Pt, (b) little overlap with items on several psychotic scales especially Pa and Ma, and (c) little overlap with Ps but considerable overlap with the ego strength scale. The overlap forms a high percentage of the total items on a few scales; the highest being 46% with Pt. The items on the prognosis scale seem quite stable, for they are not the ones which change with shock treatment. To study further the relationship between Ps and Evs, scores on the two scales were correlated. On 63 cases, the pre-shock Ps scores correlated only .33 with the post-shock EVS scores. Thus there is only a moderate but significant tendency for prognostic and evaluative scores to agree. While favorable Ps and EVS scores seem to go together, variable EVS scores tend to accompany poor Ps scores.

Table 4. Item Overlap Between EVS and other MMPI Scales

Scale	Number of Items	Common Items	Numbered scored Same by recovered	Number scored Same by unimproved
L	15	0	0	0
F	64	0	0	0
K	30	6	6	0
Hs	33	6	0	6
D	60	15	0	15
Hy	60	11	2	9
Pd	50	13	2	11
Mf	60	2	0	2
Pa	40	4	0	4
Pt	48	22	0	22
Sc	78	10	0	10
Ma	46	3	2	1
Ps	52	6	0	6
Ego strength scale	68	13	13	0

Interpretation of EVS

The process of improvement may be elucidated by an analysis of the item content on EVS compared with Ps. Before treatment, the items which distinguished good from poor prognosis pertained mainly to interpersonal relations. That is, *before* shock treatment the patient with a good prognosis perceived possibilities of good relations with the family, on the job, socially, etc., while the reverse held true for the patient with a poor prognosis. It has been suggested[1] that the responses of the *unimproved* indicate ego-alien externally directed aggression. The answers of the good prognostic bet seemed to belie his clinical status. By design the responses of all groups to EVS are quite similar before treatment. The items which change with shock treatment for the *recovered* but not the *unimproved* are, however, quite different from the content of Ps. Thus the items on EVS are mainly symptomatic in nature referring to bodily concerns, *e.g.*, A-40, "I feel weak all over much of the time"; or vague fears, *e.g.*, G-2, "Several times a week I feel something dreadful is about to happen"; or worry, anxiety, and tension, *e.g.*, G-6, "I certainly have had more than my share of things to worry about," F-45, "I brood a good deal," and I-35, "Life is a strain for me much of the time"; and to personal inadequacy, *e.g.*, I-37, "I certainly feel useless at times." The few interpersonal items reflect a feeling of social isolation, *e.g.*, E-43, "When in a group of people, I have trouble thinking of the right things to talk about." These items, then, largely refer to intra- rather than interpersonal tensions. The striking content differences between Ps and EVS point up the intricacies in understanding the verbal behavior of clinical groups.

In so far as these verbal self portraits mirror the clinical picture, we might surmise that shock treatment has mainly produced a symptom reduction; the improved patient is less tense and anxious, less preoccupied with bodily malfunctioning, less

suspicious and somewhat more optimistic, and finally less withdrawn socially. Although he seems to have a much more positive self-evaluation, there is little indication of better interpersonal relations. This portrait of the improved patient suggests a symptomatic cure. *Perhaps, these results reflect the kind of change produced by somatic treatment.*

The unimproved cases tend to be "plus" getters, that is they give statistically infrequent responses both before and after treatment while the recovered cases tend to change. In this respect the high item overlap of EVS with Pt as well as the sharp increase in K scores[3] of recovered patients after treatment can be noted. From these diverse bits of information, it is possible that EVS also in part taps the same test taking attitude reflected in such scales as L, K, F, N, and Pt[6]. Meehl suggests that more than a mere test taking attitude is inherent in the answers to such scales; there is some reason for believing that the adequacy of the patient's ego strength and defensive structure is involved. The present results lead the writer to suggest that "plus" getting in a normal record may be a sign of realistic self criticism and good ego strength while "plus" getting in an elevated profile may indicate hostility, resignation, and inadequate ego defenses. It is possible that EVS measures one facet of ego strength, namely a decrease in internal tension reflecting a more positive self-evaluation. This factor may be a pre-condition for a fuller recovery but may or may not be accompanied by better interpersonal relations.

Test scores on EVS are simple raw scores which in practical usage would be taken only from post-shock MMPI records. We assumed that high scores indicate *changes* in test behavior compared with pre-shock scores. As the scale is now used, the magnitude of the change does not affect our interpretation. It is quite possible that a "difference" score between pre- and post-shock scores on EVS might be more meaningful. However, difference scores based upon arithmetic differences between pre- and post-shock EVS scores actually turned out to be less discriminating than the results reported in this study. This finding is understandable in view of the fact that pre-shock EVS scores are somewhat lower in the *unimproved* group than in the *recovered* group. It is possible that a more refined measure which takes into account the initial discrepancy in level between the groups might produce a difference score with equal or greater discriminating power than the simple post-shock EVS score. I am attempting to work out the technical problems involved in this transformation.

Summary and Conclusions

1. This study attempted to derive empirically a scale to evaluate response to shock therapy.
2. The resultant scale, labeled EVS, can be considered moderately valid as a tool to evaluate shock therapy. Validity was impaired mainly because a sizable number of unimproved patients obtained favorable test scores.
3. Scores on EVS after treatment were positively correlated with pre-shock scores on a prognosis scale but only to a moderate degree.
4. In the main, the content of items on EVS reflected symptoms and not interpersonal relations. This suggests that improvement with shock treatment is largely a process of symptom abatement with no real changes in interpersonal relations.

References

1. Carpenter, L. G. An experimental test of an hypothesis for predicting outcome with electroshock therapy. *J. Psychol.*, 1953, *36*, 131-135.
2. Feldman, M. J. A prognosis scale for shock therapy. *Psychol. Monogr.*, 1951, *65*, No. 10 (whole No. 327).
3. Feldman, M. J. The use of the MMPI profile for prognosis and evaluation of shock therapy. *J. consult. Psychol.*, 1952, *16*, 376-382.
4. Feldman, M. J. The effects of the size of criterion groups and the level of significance in selecting test items on the validity of tests. *Educ. psychol. Meas.*, 1953, *13*, 273-279.
5. Hathaway, S. R. and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* Minneapolis: University of Minnesota Press, 1943.
6. Meehl, P. E. and Hathaway, S. R. The K factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, 30, 525-564.

A VALIDATION OF MEASURES THAT PREDICT THE EFFICACY OF SHOCK THERAPY[1]

DONALD K. PUMROY[2] AND WILLIAM S. KOGAN
University of Maryland *VA Hospital, Seattle, Washington*

Problem

In recent years psychologists have become increasingly interested in predicting whether a mentally ill person will benefit from a particular type of treatment such as insulin and electroshock (EST). Several studies have used the Minnesota Multiphasic Personality Inventory (MMPI) to predict the response to shock therapy. Feldman[(1)] empirically derived a prognostic scale (Ps) for shock therapy from an item analysis of the MMPI's of two groups of *S*s, one of which recovered and the other which remained unimproved as the result of shock treatment. This scale was then found to discriminate between recovered and unimproved patients in an independent sample. Using the MMPI profile as a basis, Welsh[(3)] developed an anxiety index (AI) and an internalization ratio (IR). He found that the higher the AI or the IR, the greater the likelihood of patients benefiting from the different types of treatment studied.

Pearson[(2)] analyzed the pre-treatment MMPI's of male patients given EST and later rated by a psychiatrist as to their response to treatment, finding three profile signs related to a poor response to EST as follows: the T score on the "D" scale was greater than 84 or less than 55, the T score on the "Sc" scale was greater than 84, at least four of the nine clinical scales had T scores of 80 or above. He found that the greater the number of these signs on a patient's MMPI, the greater the likelihood of a poor response to EST.

The purpose of this study was to apply these measures of prognosis of EST to a new group of patients in order to provide additional data on the validity of the measures.

Procedure

The subjects consisted of all testable patients from a VA general hospital who were assigned to EST between May 1953 and November 1954. During this period, 77 patients were assigned EST, and 65 of these patients had ceased EST by the end of November 1954 receiving 879 EST's or a mean of 13.5. Twenty-three patients were testable, all male veterans age 22 to 62, mean of 37.5. They received a total of 331 EST's with a mean of 14.4.

Each subject was given the MMPI from 1 to 7 days prior to the start of his EST series. The test was given with standard instructions except that the subject was told to answer all of the items. Four indices were ascertained: (a) Feldman's Ps scale, (b) Welsh AI signs, (c) Welsh IR, and (d) Pearson's three signs. These indices, of course, were determined only after two criterion measures had been obtained. The first and primary criterion was a rating of improved or unimproved made by the patient's psychiatrist. This rating was made approximately one month after the patient terminated EST. The other criterion was the number of EST's the patient received. As the policy of the hospital was to remove a patient from EST when he showed improvement, it was assumed that a smaller number of EST's was related to improvement.

Results

The psychiatrists rated 16 patients improved and 7 unimproved. These two groups are compared as to mean scores on the Ps, AI, and IR, and the number of Pearson's signs in Table 1.

Table 1. Means on the Ps, AI, IR and Pearson's Signs

Factors	Ps	AI	IR	Signs
Unimproved	26	96.6	112.9	2.1
Improved	24.3	102.8	119.6	1.7

[1]This study was conducted at the VA Hospital, Seattle, Washington, with the cooperation of the Department of Psychiatry, University of Washington.

[2]Formerly at University of Washington.

To check the significance of these differences, t tests were used. Although none were significant, all of the measures except the Ps were in the direction predicted. Table 2 indicates the correlation between the two criterion measures (psychiatric ratings and the number of EST's) and the four prognostic measures. Product moment correlations were computed between the prognostic measures and the number of EST's; biserial correlations were used with prognostic measures and the ratings.

TABLE 2. CORRELATIONS BETWEEN PROGNOSTIC MEASURES, NUMBER OF EST'S AND RATINGS

Factors	Number of EST's	Ratings
Ps	−.04	−.16
AI	−.27	.14
IR	−.42*	.22
Signs	−.37	−.29

*Significant at the 5% level.

Only the correlation between IR and the number of EST's was significant. This is in the hypothesized direction; the higher the IR the fewer the number of EST's received. The two criterion measures were not highly correlated (r = .14). Actually, there was a slightly greater chance of being rated improved with an increase in the number of EST's administered.

DISCUSSION

In general, the results of this study did not confirm the ability of these measures to predict the efficacy of EST. The reason for this lack of confirmation may lie in two areas. The first possibility concerns the sample used in studies of this nature. Probably of most importance is the selection of the individuals who are to receive EST. This is most commonly done by a psychiatrist, who selects certain patients for EST in terms of rather subjective criteria. In some cases the whole psychiatric staff has a certain unwritten policy as to which patient will receive EST. It seems quite possible, therefore, that the cases selected for EST at one hospital may be different from those at another hospital. When this takes place the measure to predict prognosis for EST devised at one hospital will be more likely to fail in another hospital. It seems important, therefore, to validate such a measure in several different settings and to objectify psychiatric indications for giving EST. In this study only 23 out of the 77 patients who were given EST were testable. Consequently, in extrapolating from these results, only one patient out of three who was to receive EST could be evaluated on these prognostic measures of response to treatment. Another possible explanation for the discrepancy may be in the criterion of improvement where raters may have had different frames of reference. There is a great need, in psychological research of this type, for a criterion that is objective, inexpensive and can be obtained shortly after treatment.

SUMMARY

Four different MMPI measures (Feldman's Ps, Welsh's AI and IR, and Pearson's signs) purporting to predict the efficacy of EST were applied to a sample of mental patients. Twenty-three male veterans were given the MMPI before they received EST. Two criteria of improvement were used: (a) ratings by psychiatrists, and (b) number of EST's received. Only the relationship between number of EST's and the IR was significant. In general, the measures were not able to predict the efficacy of EST with this sample.

REFERENCES

1. FELDMAN, MARVIN J. A prognostic scale for shock therapy. *Psychol. Monogr.*, 1951, 65, No. 10 (whole No. 327).
2. PEARSON, JOHN S. Prediction of the response of schizophrenic patients to electro-convulsive therapy. *J. clin. Psychol.*, 1950, 6, 285-287.
3. WELSH, GEORGE S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, 16, 65-75.

AN EVALUATION OF THE GRASSI TEST FOR ORGANIC INVOLVEMENT

MICHAEL HIRT

University of Nebraska

Problem

One of the outstanding symptoms of organic patients, depending upon the location and extent of the lesion, is the loss of ability to function at the abstract level. A normal subject is capable of utilizing both the concrete and abstract approach in solving problems, whereas damaged or otherwise impaired subjects have difficulty with abstract responses which are among the first affected. It therefore has become necessary to develop sensitive psychological tests for detecting intellectual defects due to organic involvement. Yates [2] has made an excellent evaluation of the psychological tests of brain damage.

The Grassi Block Substitution Test [1] consists of a set of colored cubes similar to the Kohs blocks. Each block has one side painted yellow, one white, one red, one blue, one red and white, and one yellow and blue, the latter two sides being equally divided diagonally. Each of the 24 cubes in the test is identical in size, color, and color relationships. It is important that all of the cubes be identical to the most minute detail, not only to allow the subject the possibility of completing all of the designs, but also to avoid confusing organic patients who are often unable to overlook slight color and hue differences.

Procedure

The clinical subjects used in this study were selected from two Iowa state institutions. The hospital staffs assumed complete responsibility for selecting subjects. The need for patients whose diagnosis was unequivocal was impressed upon them. Both institutions supplied patients whose clinical histories were sufficiently complete to warrant concluding the presence or absence of organic involvement. The organics were either post-lobotomy cases, syphilitics, grand mal epileptics with years of numerous daily seizures, or alcoholics of long duration. The functional group consisted almost entirely of schizophrenics. In selecting normal subjects, the only limitation imposed was the effort to use normals of comparable age to that of the patient population. The final investigation contained data from 120 cases, ranging in age from 16 to 78 years and representing intelligence levels from moron to very superior. Prior to testing the author had no knowledge of any patient's diagnosis.

In administering the Grassi test, the subject works with four cubes. With these he is to reproduce either parts of, or the complete, stimulus design. The Grassi test consists of five designs. Each design is reproduced at four levels of increasing complexity. The simple concrete level requires the subject to reproduce only the top of the stimulus design; the simple abstract level requires the subject to again reproduce only the top of the stimulus design, but using different colors. The complex concrete level requires the subject to reproduce the entire stimulus design; while the complex abstract level requires the subject to use different colors in reproducing the entire stimulus design.

An effort was made to control the variables of age and sex. Practical considerations, however, made it impossible to match the functional, normal, and organic groups completely. Although the range of ages for the three groups was not identical, the mean ages were within two years of one another. The correlation between age and Grassi test score was not reliably significant; neither was the *t* test between the mean scores for the male and female subjects.

Results

Table 1 indicates the distribution of Grassi scores for the entire sample. Of the 40 normal subjects, not one received a score below 20. Of the subjects with known organic pathology, not one scored above 16.

TABLE 1. DISTRIBUTIONS AND PERCENTILE RANKS OF SCORES ON THE GRASSI BLOCK SUBSTITUTION TEST FOR ORGANIC, SCHIZOPHRENIC AND NORMAL GROUPS

Score	Group					
	Normal		Schizophrenic		Organic	
	f	PR	*f*	PR	*f*	PR
26.0–26.9	1	.99				
25.0–25.9	4	.98				
24.0–24.9	4	.85				
23.0–23.9	6	.78				
22.0–22.9	11	.63	5	.99		
21.0–21.9	8	.35	3	.88		
20.0–20.9	6	.15	10	.80		
19.0–19.9			4	.55		
18.0–18.9			9	.45		
17.0–17.9			2	.23		
16.5			7	.18		
16.0					3	.99
15.0–15.9					1	.93
14.0–14.9					6	.90
13.0–13.9					6	.75
12.0–12.9					5	.60
11.0–11.9					5	.48
10.0–10.9					5	.35
9.0– 9.9					3	.23
8.0– 8.9					0	.15
7.0– 7.9					4	.15
6.0– 6.9					1	.05
5.0– 5.9					0	.03
4.0– 4.9					0	.03
3.0– 3.9					0	.03
2.0– 2.9					0	.03
1.0– 1.9					1	.03

The normals made the greatest number of errors at the complex abstract level. The same pattern of errors seems to repeat itself in the functional group, although there is also a marked increase in the number of errors made on the simple abstract level of the more complicated designs. The deterioration accompanying brain damage is exemplified to its fullest extent in the organic group. Abstract behavior, which is one of the first functions affected, is most noticeably absent in the performance of organic subjects. Accordingly, they committed the largest number of errors at the simple and complex abstract levels.

Since the Grassi test measures the ability to function at the abstract level, and since the ability to abstract is influenced by intelligence, the author considered it pertinent to make some effort at equating the functional, organic, and normal elements in regard to this factor. Accordingly, measures of intelligence were obtained for all the subjects, and these were correlated with Grassi test scores. A coefficient of .484 was obtained.

Grassi[1, p. 61], using the test-retest method, reported a reliability coefficient of .85 for the Grassi test. Since the validity of the Grassi test was found to be so high, it was considered adequate to accept the reliability reported by Grassi.

DISCUSSION

A subject's diagnosis should be made on the basis of an analysis of his test behavior and intellectual level. Disregarding any one of these factors will tend to lower the reliability and validity of the diagnosis. If a subject is classified on the basis of test score alone, the clinician might easily be misled as to the patient's actual level of performance. For example, a score of 18 would not be considered indicative of impairment in an individual of borderline intelligence. On the other hand, a score of 18, if obtained by a subject of superior intelligence, would be strongly suggestive of moderate impairment. Also worthy of consideration is the fact that a

score of 12 for a subject of superior intelligence is indicative of more extensive organic deterioration than the same score for a dull subject.

Grassi has suggested a score of 20 as the cutting point between organics and non-organics. The results of this study, however, tend to indicate a score of 16 as more appropriate.

Summary

One hundred twenty subjects, equally divided into organic, functional, and normal groups, were tested with the Grassi Block Substitution Test. All of the normal subjects scored above 20 and all of the organics fell below this score, suggested by Grassi as the cutting point between normal and brain damaged subjects. Although all of the funtional subjects scored higher than the organics, approximately half of them fell below 20, which, according to Grassi, would suggest moderate organic impairment. The results of this study support the general conclusion that the Grassi test should take its place among other diagnostic instruments.

References

1. Grassi, J. R. *The Grassi block substitution test for measuring organic brain pathology.* Springfield, Ill.: Charles C. Thomas, 1953.
2. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, 51, 359-377.

AN EVALUATION OF THE "ORGANIC SIGNS" IN THE H-T-P DRAWINGS*

VYTAUTAS J. BIELIAUSKAS AND SANDRA L. KIRKHAM[1]

Richmond Professional Institute

Introduction

Conspicuous changes in the performance of organic brain injured cases have been noticed since time immemorial by popular observers and literary writers. Authorities on aphasia such as Goldstein [4], Head [6], and Riese [11] have reported specific changes in verbal communication. In full agreement with these authors, Jaspers [7] went a step further, suggesting a definite change in all the performance areas of a person with cortical disturbances. Basing his hypothesis upon various reports of experimental observers, Jaspers [7, p. 244-245] suggested that drawings by organic persons must show specific characteristics, *i.e.*, omissions, exaggerations, and signs of loss of Gestalt. The research and theories advanced by more recent authors [1, 2, 9, 13] primarily concerned with the diagnostic aspects of drawings, imply that there are some differentiating characteristics in the drawings of organic patients from those of normals.

Buck [2, 3] and Jolles [8] have advanced hypotheses concerning specific signs suggesting organicity in the House-Tree-Person (later referred to as H-T-P) drawings. Some aspects of these hypotheses have been investigated by Michal-Smith [10], whose research was unable to give definitive support to Buck's and Jolles' theories, but gave some tentative indications that the trends of the findings tended to support

*Grateful appreciation is expressed herewith to Dr. W. I. Prichard, Superintendent, and Miss Hannah S. Davis, Chief Psychologist, Lynchburg Training School and Hospital, Colony, Virginia, for their assistance in securing the sample for this study, and to Dr. Nicholas M. Vincent for his advice on statistics.

[1]The co-author is now with the Department of Psychiatry and Neurology, The Bowman Gray School of Medicine, Wake Forest College, Winston-Salem, North Carolina.

Buck's assumptions. Therefore, the present study was designed (1) to restudy some findings of Michal-Smith in order to reach more conclusive evidence as to the validity of H-T-P "organic signs", and (2) to test additional "organic signs."

Subjects

Forty subjects were selected from the population of the Lynchburg Training School and Hospital. The staff diagnosis made on the basis of social history, medical history, and psychological evaluation was used as criterion for distinction between organic and non-organic subjects. Only cases diagnosed as having some definite structural damage of the brain tissue were included in the organic group. The non-organic group was selected from those classified as either familial mental deficiency or without mental deficiency. All whose diagnoses included suggestions of other disorders, *i.e.*, psychosis, neurosis, or character disorder, were excluded from the sample. Because of rigid selection, only twenty subjects were obtained which could be classified as organic. Twenty non-organic subjects were used in the control group. The groups were matched for: sex, age, months in institution when tested, form of Wechsler given, and the IQ's. The results of the matching are shown in Table 1.

Table 1. Matching of Organics and Non-Organics on Sex, Age When Tested, Length in Institution, Form of Wechsler Given, and IQ Scores

Matching Variables	Organics	Non-Organics
Sex		
Males	12	12
Females	8	8
Age when tested		
Range	14 - 33	15 - 32
Mean	22	22.65
Standard deviation	6.6	6.2
Months in institution when tested		
Range	1 - 193	1 - 235
Mean	62.95	64.30
Standard deviation	45.7	63.7
Form of Wechsler given		
Form I	14	13
Form II	0	1
WAIS	5	5
WISC	1	1
Verbal IQ		
Range	51 - 101	52 - 100
Mean	68.1	67.7
Standard deviation	11.8	11.6
Performance IQ		
Range	37 - 104	42 - 102
Mean	63.9	64.0
Standard deviation	18.2	15.2
Full Scale IQ		
Range	45 - 94	45 - 88
Mean	62.9	63.5
Standard deviation	14.3	12.7

The H-T-P drawings of each subject were scored according to eighteen objective criteria suggested by Jolles[8] and Buck[2, 3] as shown in Table 2. Only signs which could be evaluated (measured) objectively were included. All other signs mentioned by Buck and Jolles which in any way required subjective interpretation on the part of the scorer were excluded.

A Chi-square and Phi coefficient were computed for each item to determine if there were any significant differences. Pearson's *r*'s were obtained from the Phi coefficients to account for possible continuity of variables[5, p. 331]. A *t* test was run to determine the differences between the mean numbers of the scores on objective criteria of both groups.

TABLE 2. EIGHTEEN OBJECTIVE ITEMS INDICATING ORGANICITY IN H-T-P DRAWINGS FROM BUCK[2, 3] AND JOLLES[8]

House	Tree	Person
1. Details, irrevelant: horizontal line separating first and second stories.	8. Trunk, one-dimensional, with one-dimensional branches that do not form a system.	13. Head, large in relation to trunk (H:T less than 1:1).
2. Foundations: pillars abnormally high (height of pillars one-fourth height of house or higher).	9. Dangling tree roots (roots above the ground line).	14. Fail to join . . . arm or legs to trunk.
3. Steps leading to blank wall.	10. One or two dimensional trunk and one or two dimensional branches some of which are not attached to trunk.	15. Full face, except perhaps, for the head.
4. Walls, disconnected.	11. Tree drawn as falling (angle from trunk to base of paper less than 75 degrees).	16. One-dimensional neck.
5. "Facade" drawn.	12. Omits one or more of the "essential" details. Essential details for tree are a trunk and one branch.	17. Person drawn as falling (angle of trunk to base of paper less than 75 degrees).
6. House drawn as toppling (angle of 75 degrees or less —measuring from base line of house to axis of paper).		18. Omits one or more of the "essential" details. Essential details for a person are a head, two eyes, two ears, a nose, and a mouth, and a trunk with two arms and two legs.
7. Omits one or more of the "essential" details. Essential details for house are a roof, a wall, a door, a window, and a chimney.		

RESULTS AND DISCUSSION

The Chi-squares, Phi coefficients, Pearsonian *r*'s, and *t* tests are presented in Table 3. None of the eighteen items shows a Chi-square significant at the 5 per cent level. No significance can be deduced on the basis of either Phi coefficients or the converted Pearsonian *r*'s. Neither does the Phi coefficient for both groups indicate differences sufficiently significant for individual prediction. Likewise, the *t* test values are not significant.

TABLE 3. NUMBER OF ORGANIC AND NON-ORGANIC SUBJECTS SHOWING THE CHARACTERISTICS OF THE 18 OBJECTIVE CRITERIA OF ORGANICITY WITH CHI-SQUARE, PHI COEFFICIENT, PEARSON r, AND t TEST

Item No.	Organic	Non-Organic	X^2	ϕ	r_ϕ	*t*
1	2	2	0	0	0	
2	0	0	0	0	0	
3	1	1	0	0	0	
4	2	3	0	.075	.127	
5	12	9	.4	.150	.234	
6	0	0	0	0	0	
7	9	6	.426	.131	.221	
8	3	1	.276	.166	.298	
9	5	2	.663	.197	.322	
10	5	5	0	0	0	
11	1	1	0	0	0	
12	0	1	0	−.150	−.272	
13	6	6	0	0	0	
14	2	6	1.404	−.250	−.416	
15	4	1	.914	.226	.396	
16	0	0	0	0	0	
17	1	4	.914	−.226	−.396	
18	11	10	0	.050	.078	
Total	64	58		.022	.036	.55

Since the statistical values obtained on the objective scoring criteria do not show significant findings, it appears that the signs postulated by Buck and Jolles as

discriminators of organic brain damage used in this study were inadequate to discriminate between the drawings of this population. The highest correlation was negative (–.42) on the item referring to failure in joining of the arms or legs to the trunk. The drawing of the person as toppling was also negatively correlated (–.40) with organicity, while the item dealing with drawing of a person in full face, except for head, was positively correlated (.40) with organicity. Since all of these correlations are far from significant for individual predictions, and only very few are approaching values required for group prediction, it seems that the items examined in this study are not valid discriminators of the organicity.

There are two objections which could be raised against this conclusion. First, it may be said that most mental defectives might be conceived as organics, and since our study included some subjects classified as familial mental deficiency, the comparison would not tend to distinguish adequately between the two groups. Should the idea that most mental defectives are also organics prove to be correct, then it would seem that there would be a trend indicating higher frequency of items appearing in the drawings of both groups. However, such a trend cannot be perceived in the results obtained. Only two items exceeded half the possible score for each group (items 5 and 18), while sixteen other items fell considerably below.

Second, the smallness of the sample is well recognized. However, it should be pointed out that twenty subjects was the maximum sample which could be obtained from a population of 2300, in applying the rigid criteria of matching and diagnosis. Therefore, the sample could be considered as highly representative. Of course, it would be desirable to repeat this study on a larger sample, if such a sample meeting all criteria could be obtained.

Nevertheless, it seems that it is reasonable to assume that at least a cautious attitude is indicated in using the so-called "organic signs" in H-T-P drawings as organicity discriminators, as long as the support for their validity rests on such tenuous grounds as the Michal-Smith study and the present findings. A recent study by Silverstein and Robinson [12] of drawings of poliomyelitic children done for another purpose but in which H-T-P criteria were used did not bear out any signs of significant frequency, giving again some support to our findings with adults.

Summary and Conclusions

Sets of H-T-P drawings were obtained from forty white institutionalized subjects including 20 organic and 20 non-organic cases. The subjects were matched for sex, age, months in institution when tested, form of Wechsler given, and the IQ's. The drawings were examined for 18 objective criteria claimed to be indicative of organicity. Chi-squares, Phi coefficients and Pearsonian r's were computed on each item and a t test on the difference between the mean scores of groups was run. None of the statistical results showed significant differences. Therefore, it was concluded that the eighteen criteria claimed to be indicative of organicity did not adequately discriminate either individually or together the organics in the population of this study. Although a larger sample may give more conclusive evidence, it was suggested that extreme caution be exercised in using these signs as organicity indicators in H-T-P drawings.

References

1. Bender, Lauretta. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Assoc.*, 1938, Research Monograph No. 3.
2. Buck, J. N. *Administration and interpretation of the H-T-P test: proceedings of the H-T-P workshop held at Veterans Administration Hospital, Richmond 19, Virginia, March 31, April 1, 2, 1950.* (Better known as: *The Richmond Proceedings*). California: Western Psychological Services, 1950. (Mimeographed)
3. Buck, J. N. Signs of H-T-P suggesting organicity. Unpublished manuscript.
4. Goldstein, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.
5. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw Hill, 1956.
6. Head, H. *Aphasia and kindred disorders of speech.* New York: Macmillan, 1926.

7. JASPERS, K. *Allgemeine psychopathologie*. Berlin and Heidelberg: Springer, 1946.
8. JOLLES, I. *A catalogue for the qualitative interpretation of the H-T-P*. Beverly Hills, California: Western Psychological Services, 1952.
9. MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: C. C. Thomas, 1949.
10. MICHAL-SMITH, H. The identification of pathological cerebral function through the H-T-P technique. *J. clin. Psychol.*, 1953, 9, 293.
11. RIESE, W. Hughlings Jackson's doctrine of aphasia and its significance today. *J. nerv. & ment. Dis.*, 1955, 122, 1-13.
12. SILVERSTEIN, A. B. and ROBINSON, H. A. The representation of orthopedic disability in children's figure drawings. *J. consult. Psychol.*, 1956, 5, 333-341.
13. WARTEGG, E. Gestaltung und Charakter. Ztschr. ang. Psychol., Beih. 84, 1939.

HUMAN FIGURE DRAWINGS BY YOUNGER AND OLDER ADULTS

IRVING LORGE, JACOB TUCKMAN AND MICHAEL B. DUNN

Teachers College, Columbia University

PROBLEM

Dunn[1] has demonstrated that drawings of the self by boys and girls, 4 to 18 years of age, show a progression in maturity of representation, from an undifferentiated drawing at age four or five, to a well-integrated human figure at age fourteen. Drawings of post adolescents seem to be at the same level as those of 14 year olds. Dunn suggested that human figure drawings may be used as evidence of psychic intactness of the self in that the drawings are positively correlated with age, grade, intelligence and aspects of adjustment. The aspect of adjustment appraised was the distinction between the drawings of deviants (feeble-minded, neurotic, psychotic, and brain damaged) and those of presumed normals. The drawings of deviants were found significantly less mature than those of normals of the same chronological age.

Presumably, if people suffer physiological and psychological deficits as they grow older, such changes may be reflected in their drawings of the self. The purpose of this study, therefore, is to test the hypothesis that the drawings of older people show increasing loss of intactness. Drawings of sophisticated younger adults where physiological and psychological deficit should be relatively minor will be compared with those of older adults where the probability of such deficit should be greater.

SAMPLE AND METHOD

The younger adults were seventy-five students (average age 32) enrolled in a graduate course on the psychology of the adult; the older adults (average age 73) were 104 individuals, some living in an institution for the aged and some participating in recreational day centers for old people in New York City.

> "Draw a picture of yourself, standing, facing front view as if you were looking at yourself in a full-length mirror. We are not the least bit concerned with whether or not you can draw like an artist. We are not measuring art or art ability. Don't think too much about it, but just draw. Be at ease and whatever you draw will be satisfactory."

For the younger adults and for most of the older adults, the drawings were completed in group sessions. The younger adults were given a time limit of six minutes; the older subjects had no time limit. Just a few needed more than fifteen minutes to complete the drawing.

Results

The first appraisal of the drawings rated them on the 20-point scales, one for boys and the other for girls, made by Lorge and Dunn [1] to evaluate holistically the maturity of the representation of drawings by children and adolescents. The mean score for the drawings by graduate students and by older subjects sharply differentiates between the younger and older adults. For younger adults, the mean scale value for men is 14.1, and for women, 14.0, which correspond to the point at which a leveling takes place among children and adolescents. For older adults, the mean scale value for men is 7.7, which corresponds to the average drawings by nine-year old boys; for women, 7.3, which corresponds to the average drawings by seven and one-half year old girls.

The method of appraisal by the children's scale fails to give consideration to the qualitative differences between children's drawings and those by younger and older adults. For instance, in contrast to the drawings of children, those of graduate students are of "more mature" persons, *e.g.*, the human figure looks fuller, the face looks older, wrinkles show, clothes are more adult, glasses are shown, etc. The qualitative differences between children's drawings and those by older adults are even greater. When a recognizable human figure is drawn by an older adult, the drawing looks like a person beyond the age of 50, *i.e.*, the face looks older, more wrinkles are shown, *etc.*

Superficially, some aspects of the drawings of older adults and of very young children are similar. The drawing of a four year old, as may be expected, is so immature that the sex of the respondent cannot be determined. Of the drawings by older men and women, 24 percent are as poor or poorer than those of the average four year old, and two-thirds of these cannot be identified for sex of the respondent. Yet, these adult drawings of the self do not look like those drawn by children.

Since the Lorge-Dunn scale was not a wholly satisfactory way of appraising the drawings by adults, seven criteria were developed for comparing the drawings of younger and older adults: (1) identification of the sex of the respondent, (2) intactness: no omission of a major part of the body, *e.g.*, legs, arms, hands, feet or torso, (3) integration, *i.e.*, evidence of body gestalt, (4) depth (tridimensionality or the impression of solidity) rather than flatness, (5) good proportion of the body parts, (6) absence of bizarreness, and (7) evidence of adequate motor coordination.

Appraising the adult drawings by these seven "either/or" criteria gave evidence of significant differences between the younger and the older adults. For younger adults, the sex of every respondent could be identified; for the older adults, sex could not be identified for 18 percent and for 13 percent could not be identified correctly. For the younger adults, 85 percent of the drawings were intact compared with 25 percent for the older adults. Moreover, there were qualitative differences between young and old. The younger adults omitted such parts of the body as hands or feet, but the older adults omitted not only hands and feet, but also arms, legs, and torso. In a number of cases, the drawings by older adults were just shells, *i.e.*, just a head, supported by legs with the entire torso missing. For the younger adults, all drawings gave evidence of a body gestalt, compared with 45 percent of older adults. For the younger adults, 77 percent of the drawings gave evidence of depth of the human figure, compared with 18 percent of the older adults. For the younger adults, 97 percent of the drawings showed good proportion of the body parts, but only 45 percent of the drawings of older adults did so. For the younger adults, 97 percent of the drawings were not bizarre, compared with 62 percent for the older adults. For the younger adults, no drawing suggested evidence of poor motor coordination, but 13 percent of the older adults suggested such evidence.

Undoubtedly, the different criteria are interrelated, *i.e.*, an intact drawing tended to show body gestalt, good proportion of the body parts, depth, absence of bizarreness, adequate motor coordination, and made possible estimates of the sex of the subject, while a drawing lacking intactness tended to show inadequate gestalt, poor proportionality, *etc.* Using a numerical weight of one for each of the seven

criteria, two-thirds of the drawings by younger adults received a maximum score of seven; none, a score of less than five. By contrast, only one percent of the drawings by older adults received a score of seven and 58 percent a score of zero to four.

The two methods of appraisal overlap insofar as the Lorge-Dunn scale gave holistic consideration to the seven criteria. The correlation between the seven criteria score and the scale value by the Lorge-Dunn scale is .75 for older women, .93 for older men, .25 for younger women, and .45 for younger men. The correlations for older men and women and younger men are significantly different from zero. The correlation between score for criteria and scale value underestimates the relation for younger adults because of the marked restriction in range, *i.e.*, two-thirds were rated favorably on every criterion. The relationship between the two appraisals tends to be closer for men than for women although significant only for the older adults. Although the relation between the two appraisals is close, separate scales for younger and for older adults, similar to the Lorge-Dunn approach for children, seem desirable to account for the qualitatitive differences in adult drawings.

Discussion

The sample of younger adults and of older adults cannot be considered to be representative either of younger adults or of older adults in the population. Many of the differences found in this study may be a function of the method of getting the cooperation of the individuals who participated in the experiment, as well as where they happened to be when such requests were made. At best, the evidence does suggest that older adults, whether in homes for the aged or ambulatory participants in a recreational program for the aged, do make drawings of the self which differ markedly from those of graduate students and from children in general. While it is true that the correlation between scale values established on children's drawings and criteria of proportionality, intactness, and so on, do correlate significantly, correlations suggest that something more, or rather, less, is in the drawings of older adults.

For younger adults, however, the drawings are so homogeneous, in general, both with reference to scale value and with reference to the criteria, that the data must be considered as evidences of the fact that there is a leveling off in adolescence which is maintained up to the later 30's. The use of drawings of the self as a further part of the evaluation of the adult personality should lead to significant contribution to the aging process.

Summary

A comparison of the drawings of the self by graduate students (average age 32) with those by institutionalized and noninstitutionalized older adults (average age 73) supports the hypothesis that the drawings of older people show increasing loss of intactness. In contrast to the human figure drawings by younger adults, the drawings of older people are characterized by incompleteness, lack of integration, flatness, lack of proportion, bizarreness, evidence of inadequate motor coordination, and difficulty in identifying the sex of the respondent. In contrast to the drawings by older adults, those by younger adults are quite homogeneous. It must be recognized that the samples in this study are not representative of younger adults or of older adults in the population.

Reference

1. Dunn, Michael B., Global evaluation of children's drawings of "person" and "self". Ph.D. dissertation, Teachers College, Columbia University, 1955.

MANIFEST CONTENT AND INTERPRETIVE MEANING OF VERBAL INTELLIGENCE TEST RESPONSES

NORMAN TALLENT

Veterans Administration Center, Kecoughtan, Virginia

Problem

How valid is content interpretation of responses to verbal intelligence test items as an indicator of overt behavioral trends? This study attempts to validate Rapaport's finding [1] that subjects who indicate they would yell "fire" in response to Wechsler's [4] Comprehension item 2—the "theatre item"—show impulsive trends in their everyday behavior. In addition, other responses which appeared to suggest that the subject is impulsive, or that he shows tendencies to impulsive behavior, were examined against evaluations of behavioral control.

Procedure

The sample consisted of 198 boys who constituted the ninth grade of a public school system. Their behavioral control was appraised in terms of averaged ratings by their teachers on the Behavioral Control Rating Schedule (BCRS). [2, 3] which provides ratings of behavioral control in terms of a number of postulated component variables. An attempt was made to secure five ratings for each subject, but practical considerations reduced the number obtained to an average of 4.33. On the basis of these averaged scores of the BCRS, 120 boys taken from the larger group were assigned to one of three experimental groups of 40 each, labelled "Well-Controlled Group", "Impulsive Group", and "Average Group". All of these subjects were then tested with the Wechsler-Bellevue Intelligence Scale, Form I.

Results and Discussion

A number of the responses given might be indicative of tendencies to impulsive behavior in accordance with the hypothesis that those who say they would yell "fire" are impulsive. Thus, three subjects of the Well-Controlled group, and two subjects of the Average group gave such a response. None of the Impulsive group responded in this fashion. The hypothesis that subjects who say they would yell "fire" are impulsive is thus unsupported when impulsivity is defined in terms of low behavioral control as rated on the BCRS.

In addition to testing this hypothesis, it seemed worthwhile to determine if any classification of responses could be established which might be indicative of behavioral undercontrol. Thus all responses which mentioned some form of overt impulsive action were tabulated. Similarly, responses which suggested that the subject might have impulsive tendencies were listed. Altogether, five categories of responses having a possible relationship to behavioral control were identified as follows:

1. Yell "fire." Also shout, scream, etc.
2. Letting others know (except manager or ushers), or warning people, but no specific mention of yelling.
3. Other overt impulsive actions, e.g. running for fire extinguisher, the manager, etc.
4. Denial of yelling "fire," possibly because of consequences.
5. Mention of concern over panic, stampede, riot, danger alarming people, etc.

The rationale for this classification of possible indicators of impulsivity would probably be acceptable to most clinical psychologists. Warning others of the fire seems to be an unthinking response, perhaps almost on a par with yelling out "fire." Category 3, "other overt impulsive actions," includes such behavior as running, which could precipitate panic. Category 4, "denial of yelling 'fire,' " is included to check a variation of the main hypothesis; that is, the responses included in this category *might* indicate impulsive tendencies which are under control. If so, the question

is raised as to which of the experimental groups favor such a response. Being concerned with panic, Category 5, might be regarded similarly.

To determine whether there is a relationship between the sort of test responses listed above and behavioral control, all responses of the subjects were classified by two psychologists as belonging to one or more of the above five categories, or into a sixth category which included all responses which could not be classified into either of the above.[1] (Both psychologists showed complete agreement in classification, with the exception of one point of difference which was readily resolved.) The categories were then further combined, as follows. Categories 1, 2, and 3 were combined on the basis of including responses which mention overt impulsive actions, and categories 4 and 5 formed a new composite category to include responses which might be indicative of impulsive trends. The group of responses which could not be included in either of the above then formed another category. Using a chi square test of independence, no relationship was found between these categories and level of behavioral control ($P = <.30 > .20$). It thus appears that considerable caution is needed in predicting actual behavior from verbal responses to hypothetical problems.

Summary

The presence of a possible relationship between certain intelligence test responses suggesting that the subject is impulsive and ratings of behavioral control was investigated. The findings were negative. Considerable caution is needed in predicting actual behavior from verbal responses to hypothetical problems.

References

1. Rapaport, D., Gill, M. and Schafer, R. *Diagnostic psychological testing.* Vol. I. Chicago: Year Book Publishers, 1945.
2. Tallent, N. The relationship of behavioral control ratings to intelligence. Unpublished doctor's dissertation, Columbia University, 1954.
3. Tallent, N. Behavioral control and intellectual achievement of secondary school boys. *J. educ. Psychol.*, 1956, 47, 490-503.
4. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

[1]Interested readers may secure the detailed statistical data by writing the author.

RORSCHACH *M* RESPONSES AND INTELLIGENCE

ROBERT SOMMER

Larned State Hospital

Problem

The relationship between Rorschach human movement responses and intelligence has been the subject of considerable research. Positive and significant correlations have consistently been found, attesting to the genuineness of the relationship [1, 2, 3]. However, there exists the question of the meaning of this correlation as Rorschach M is well confounded with Rorschach H (human content). It can be asked whether it is the use of human content rather than the attribution of movement that relates to intelligence. Previous investigators have also found that M correlates highly with number of responses, Rorschach R, and that R correlates significantly with intelligence [2]. If this is the case there exists the question of whether the correlation between M and IQ is an artifact of the correlations of R with IQ and of M with R. There is also the problem of whether M responses from subjects of varying IQ are distinguishable on bases other than quantity.

Procedure

From the files of the Psychology Department of Southeast Louisiana Hospital, the records were selected of all patients for whom Rorschach protocols and Wechsler-

Bellevue records were available.[1] This yielded a population of 123 cases, 77 males and 46 females, with a mean age of 32.2, and from all diagnostic categories.

In the second stage of the study, from the files of the Psychology Departments of Southeast Louisiana Hospital and Gulfport (Miss.) V. A. Hospital, the records were selected of male patients who had been given both the Wechsler-Bellevue verbal scale and the Rorschach, and fitted into any one of nine categories based on number of M responses(2, 3, or 4) and IQ (ca. 80, ca. 100, ca. 120). This provided a nine-celled matrix with each IQ level and number of M responses represented. Every effort was made to secure representative IQ estimates by eliminating subjects whose Wechsler disclosed considerable scatter. The M responses were typed on cards and placed in groups of three depending upon the number of M responses.

Three groups of six judges each were utilized: senior psychologists, each with a Ph.D. in clinical psychology; psychology interns; hospital secretaries. Each judge was asked to rate three series of cards: (a) The verbatim M responses and inquiries as found in the original protocols, (b) in another set of cards (different responses) all obvious cues of vocabulary and grammar were removed without the judges' knowledge, *e.g.*, the responses "two harlequins in a dance" was changed to "two clowns in a dance" and the response "two guys is running" was made to read "two men are running", and (c) in the third series (again different responses) all cues were removed except the card to which the response had been given, the type of movement, and the gender of the person moving. If the gender were unknown (*e.g.*, clowns, elves) the male gender is used.

For the verbatim and language equated series, the senior psychologists and interns were told that these were M responses from Ss at IQ levels 80, 100, and 120, and that they were to rank the three cards in each group in terms of the probable IQ of the respondent. The secretaries were told simply that these were statements made by people and they should rank them from high to low in terms of what they would estimate the person's IQ to be. With the "verb only" series, all groups were told in what ways the responses had been impoverished.

As it was desirable to have each judge rate responses at each level of response structure yet not feasible to have him rate the same responses at more than one level, three replications of the design were carried out, each with a different set of responses. This permitted each judge to rank a verbatim series, a (different) language equated series, and a (different) "verb only" series. The order in which the series were given to the judge was counterbalanced.

Results

The first correlation made was between number of human movement responses and verbal IQ. This proved to be .371 ($p < .01$)[2] and confirmed the relationship found in previous studies. The correlation between number of human responses and IQ is .330 ($p < .01$). In Table 1, R is found to correlate .291 with IQ ($p < .01$).

This necessitates the use of partial correlation techniques to assess the relationship between M and IQ with the other variables held constant. These coefficients also appear in Table 1 and are quite informative. With the effects of R taken into account, both M and H continue to correlate significantly with IQ. Also of interest is the disentangling of the H component from the M score. When H is held constant, we find M just barely correlating with IQ, while when M is held constant, H loses any relationship to IQ.

With the verbatim M responses, all three groups of judges scored significantly above chance expectancy in their IQ rankings (probable errors of 4.05, 4.91, and 2.89 respectively). With the language-equated series, the interns and the senior psychologists reached the .05 probability level, while the secretaries were somewhat below

[1]The wisdom of using a "randomly selected" psychiatric population can be questioned. At most however, this fact serves to limit generalization of the results to other patient populations.

[2]All correlations are Pearson r's.

TABLE 1. CORRELATIONS AND PARTIAL CORRELATIONS BETWEEN M, H, R, AND IQ

Variables Correlated	Partialed Out	r	p
M IQ		.371	.01
H IQ		.330	.01
R IQ		.291	.01
M H		.857	.01
M R		.505	.01
H R		.497	.01
M IQ	R	.271	.01
M IQ	H	.181	.05
R IQ	M	.130	N.S.
R IQ	H	.155	N.S.
H IQ	R	.223	.05
H IQ	M	.025	N.S.

chance expectancy. With the "verb only" series, none of the groups scored significantly above chance expectancy.

A summary of the overall analysis of variance of the data is presented in Table 2. It shows that the correctness of judgments does not vary with the number of M responses per subject. The differences in performance between senior psychologists, interns, and secretaries did not attain statistical significance. The only significant variable is response structure (whether the cues of vocabulary and grammar are deleted or not). Subsequent t tests disclosed that the rankings of the verbatim material were significantly better than rankings on either the "language equated" or the "verb only " series. The difference between the latter two series is not significant.

TABLE 2. ANALYSIS OF VARIANCE OF CORRECT RANKINGS

Source*	SS	df	S^2	F	p
Training level (intern, secretary, etc.)	6.31	2	3.16	2.21	n.s.
Number of M Responses	4.79	2	2.40	2.07	n.s.
Structure (verbatim, language equated, etc.)	24.98	2	12.49	10.76	.001

*None of the interactions were statistically significant.

DISCUSSION

The correlations add further evidence for the genuineness of the relationship between Rorschach M and intelligence. They also show that the correlation is not simply an artifact of the relationships between H and IQ or R and IQ. The rankings as to IQ attest to the importance of cues such as grammar and vocabulary level in estimations of intelligence. When presented with verbatim M responses, all groups scored significantly above chance in their rankings on intelligence. When the cues of vocabulary and poor grammar were removed, the psychologists still exceeded chance expectancy while the secretaries did not. This indicates that there are cues to intelligence in Rorschach M responses other than vocabulary level and grammar. The fact that none of the groups exceeded chance expectancy with the "verb only" series tells us that these cues are not the type of movement involved or the card to which the response is given. This still leaves the response phraseology (other than good or poor grammar) and the subject's elaboration of his response.

Although the present study has shown that the responses of subjects at different IQ levels can be distinguished by using cues of vocabulary and grammar, it leaves unanswered the specific question of whether "movement quality" varies in some way with intelligence. To many clinicians, the subject of IQ 70 or 80 who gives four or even six M responses has always been an enigma. Further direct comparison of these M responses with those from subjects at other intelligence levels should help provide some of the answers.

Summary

To determine whether the relationship between M and IQ would appear with a psychiatric population and with the effects of R and H held constant, correlations and partial correlations were made between scores on the Wechsler Verbal Scale and number of Rorschach M responses for 123 psychiatric patients. The correlation between M and IQ was supported with both H and R held constant.

To determine whether M responses of Ss at different IQ levels are qualitatively different, M responses from psychiatric patients at IQ levels of 80, 100, and 120 were ranked by three groups of judges (senior psychologists, interns, and secretaries) as to intelligence. In one series of responses, all obvious cues of vocabulary level and poor grammar were removed. The results disclosed that for the verbatim M responses all groups of judges were able to exceed chance expectancy. When the obvious cues were removed, only the psychologists were able to exceed chance expectancy.

References

1. Altus, W. D. and Altus, G. T. Rorschach movement variables and intelligence. *J. abnorm. soc. Psychol.*, 1952, 47, 531-533.
2. Spiegelman, M. Rorschach form-level, intellectual functioning and potential. *J. proj. Tech.*, 1956, 20, 335-343.
3. Tucker, J. E. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.*, 1950, 14, 283-286.

GLOBAL EVALUATION OF RORSCHACH PERFORMANCE VERSUS SCORES: SEX DIFFERENCES IN RORSCHACH PERFORMANCE*

T. W. RICHARDS AND DAVID C. MURRAY

Louisiana State University *Gulfport Division, VA Center, Biloxi, Miss.*

Problem

Studies of the Rorschach test and its utility in discriminating between various types of personalities have shown that the customary formal scores are of little use. Knopf[3], for example, finds that patients classified as schizophrenic, neurotic, or psychopathic personalities are very similar in their production of location, determinants, etc. In the present study, the aim was to determine whether important differences between personalities may be elicited with the Rorschach test by means of clues other than the conventional scores. Felzer[2], reviewing the literature, shows that there is little if any difference between males and females when usual scoring categories are used for comparison. We asked ourselves this question: using the rather simple criterion of belonging to one sex or the other, is it possible for a judge to distinguish males from females, utilizing any signs or impressions he wishes to use?

Procedure

The records of thirty subjects were given (without sex identification) to each of several judges, who were asked simply to sort the records as masculine or feminine. These thirty records were selected from protocols obtained by the junior author[4] from one hundred and two students in a course in elementary child psychology at Northwestern University, twenty-four men and seventy-eight women. On a random basis, fifteen records of each sex were selected for the sorting procedure of this study.

All references in the records that might identify the sex of the testee were eliminated. Nine judges* were used; each was given the thirty protocols with a summary profile of Beck scores. Each was told that the subjects were college students enrolled in a course in child psychology, and that half the records were given by males and

*We wish to express our thanks for their cooperation to Thomas Atkins, Herdis Deabler, Arvord Belden, Bernard Eingold, Miles Hardy, Lillian Hopkins, Elizabeth Price, and Norma Schlicht.

half by females. He was asked merely to indicate which records were those of men and which those of women. He was told that he could use any procedure at all in making his decision.

Since the psychologist's experience is an important variable, judges were selected from three levels of sophistication and experience with the Rorschach test: three were Ph. D. clinical psychologists, three trainees working for the Ph. D., and three were non-psychologists (two nurses and a secretary).

Results

Differences between sexes in objective Rorschach Scores. Chi Square and *t* for differences between means were used to test significance of differences between men and women for the following objective scores: M, FM, m, F, FC, CF, C, Ft, Y + YF+FY, V + FV, Sum C, R, H, Hd, H + Hd, A, Ad, A + Ad, Abst., An., Anthrop, Art, Bot, Clo, Fire, Food, Household, Landscape, Personal. Of these twenty-nine variables, only two differentiated between the sexes at a level better than chance (Table 1).

Table 1. Rorschach Factors Differentiating Males and Females at Better Than Chance Levels

Factors		Men	Women	Probability Value
M	Mean	3.1	4.7	
	Median	2	3	.02
A + Ad	Mean	15.9	13.5	
	Median	14	12	.05–.10

It may be concluded that there are no significant differences between the sexes in terms of conventional scoring categories. Thus, our groups are similar in this respect to those of Felzer[2] and others.

"Global" judgments of sex identification. The expected frequency of identification for each cell in a fourfold table by chance alone would be 7.5 (7 or 8); that is, half the men would be designated as male on the basis of chance alone. The number of judges matching a given sex at various levels of success, together with the (Yates) corrected and uncorrected P values of X^2 for such a fourfold table are shown in Table 2.

Table 2. Success of "Global" Judgments of Sex Identification

Correct Match of One Sex	Number Judges	Probability Yates corrected	Uncorrected
8 or less	2	> .99	Near .70
9	3	.30 — .50	.20 — .30
10	1	.10 — .20	.05 — .10
11	2	.02 — .05	.01 — .02
12	1	< .01	< .01

The mean value of correct matchings was 9.44 out of 15 possible; corrected and uncorrected P values for 9.44 are about .30, and .10 - .20. When only the three trained psychologists are used, the mean value of correct matchings was 11.33. P values, corrected and uncorrected for this degree of success were < .01 in each instance.

Individual differences among subjects in their evaluation. The mean number of judges who gave proper-sex evaluations for each sex were of course identical; of nine judges in each case, 5.7 judged appropriately. There was a tendency for men subjects to spread over a greater range (two men were considered male by all judges, one man by only one judge; no women were considered female by all judges, while one woman was considered male by only one evaluator). The standard deviations of numbers of judges making proper evaluations for males and females were 2.25 and 1.93 respectively. More men were judged unanimously to be masculine than were women judged to be feminine.

Qualitative impressions of variant records. A qualitative, impressionistic examination of the eleven records identified most incorrectly and the twelve records most correctly identified suggest certain consistent patterns. Two patterns were found among the "incorrectly identified records," (at least 5 of the 9 judges identified them incorrectly). One of these, which was found in two of the female records, consisted of an isolated reference to a feminine activity which could have led to correct identification: one subject spoke of a detail as looking like "the pole I dance around every night . . . just because we do this dance," and the other said a detail looked like "sea horses . . . I cut out about 55 for a party." This isolated feminine percept was minimized by most judges, possibly because of greater emphasis on masculine percepts, such as, anatomical references, having taken biology, an airplane, percepts concerned with violence or death, or a politically oriented percept. The other pattern, found in all six of the most incorrectly identified male records, and three incorrectly identified female records, consisted in what might be called "ambi sexual" responses.

A number of percepts and verbalizations suggested femininity, and a balancing number suggested masculinity. Male indicators include masculine vocabulary and ways of talking ("that louses up the story"), technical interests in anatomy, sexual references, references to war, etc. Female indicators include artistic interests, attention to clothing, literary interests, household decorations, interest in dancing, interest in children and childish things ("all primped up", "fluffy").

Among the twelve most frequently identified male and female records (at least seven of the nine judges identified them correctly), there were several different patterns. The most frequent one, found in three female and three male records, consists of an overwhelming number of either male or female indicators, with few or none suggestive of the opposite sex. One female record and one male record had little evidence in the proper direction, but even less, or no, evidence in the wrong direction, so were usually judged correctly. In one male record, a very suggestive response (of wandering in the swamp when younger) apparently was so associated with maleness by the raters as to allow them to identify the record correctly despite some rather feminine percepts. The final three records which were mostly identified correctly seem to be rather evenly balanced between male and female indicators; they may well represent lucky guesses by some of the judges.

Since it is an observable fact that some men have more masculine characteristics than others and some women appear to be more feminine than others, it is expected that some of the errors in identification reflect real variations in the femininity or masculinity of the subjects. Thus it would seem a plausible hypothesis that the one man whose sex was misjudged by eight of the nine judges was relatively feminine and that the one woman whose sex was misjudged by seven of the nine judges was relatively masculine. Also it may be assumed that the two men correctly identified by all nine judges were masculine and the three women correctly identified by all but one of the nine judges were relatively feminine. Unfortunately, there is not sufficient information available on these subjects to test this hypothesis.

It may be noted that the selection of the subjects, college students in a class in child psychology, serves to decrease somewhat the differences between sexes which might be revealed by the Rorschach test. Thus a particular interest in children and the interests of children, which might be considered feminine, may be a poor indicator when all of the male subjects have elected to take a course in child psychology. Female college students frequently get into classes in physics, chemistry, biology, zoology, etc., and pick up some familiarity and perhaps even some interest in such things. Thus their records in that respect become more like those of the males. In general, college men might be expected to have a greater interest than men generally in art, literature, reading, etc. Thus their Rorschach responses might have more feminine indicators than would be otherwise expected.

Judgment of sex according to training in psychology. When the correct matchings for each sex were evaluated according to the level of training of the judges, it was found that the three trained clinical psychologists correctly identified a mean of 11.33 out

of 15 possible ($p < .01$, corrected for continuity). The partially trained psychologists correctly identified a mean of 9 out of 15 possible (not significant), and the untrained judges also correctly identified a mean of 9 out of 15 possible. This would suggest at least a small superiority as a result of training. However, it was noticed that some of those who did particularly well had reported taking a rather long time in making their judgments. Therefore, each judge was asked to give a rough approximation of how long he had taken to read the records and make his judgments. A rank order correlation was then computed between the judge's number of correct matchings and his own rough estimate of the time it took him to make the matchings. The resulting rank order correlation was .938. This suggests that willingness to take time to work carefully through the records was more important by far than level of training. Furthermore, it suggests that Rorschach records do contain indications of the sex of the testee, but that these indications often are not obvious to a casual or hurried observer. At our present level of experience with the Rorschach such judgments take considerable time and thought if any high level of accuracy is to be achieved.

Relationship between evaluation of men as masculine and Rorschach "signs" of homosexuality. One theory of a tendency to behave like the opposite sex is that the subject in this way is homosexual. We wanted to test this hypothesis, and decided, therefore, to correlate the degree to which subjects (men only) who were judged frequently to be feminine exhibited the Wheeler signs[5]. The fifteen men gave an average of 2.3 signs, comparable with the mean for the "normal" twenty male students examined by Davids, Joelson, and McArthur[1].

The rank difference correlation *rho* was found to be .14 between the number of times a man's record was judged female and the number of Wheeler signs.

An appraisal of Wheeler's signs, particularly as they are presented by Davids *et al*[1], suggests that these signs are deviations from the normal masculine response not so much in the direction of femininity as toward autism; there are contaminations, confabulations, drop in quality of form level, mood qualities, etc., similar to psychotic responses.

Considering these few Northwestern men students to be a valid sample, it seems appropriate to suggest that apparent femininity in men is not a regression to a less mature sexuality. It is a genuine extratension and femaleness of a healthy sort. Homosexuality in men is not femininity as much as it is a reversion or regression or arrest at an immature level of sexual development.

Summary and Conclusions

Men and women do not differ significantly in their Rorschach performance as scored formally in the usual way. When records of Rorschach performance are judged as masculine or feminine on the basis of global impression, better than chance sortings can be made by trained psychologists, particularly if adequate time is given to making the judgments. There is no relationship between the tendency for men's records to be judged feminine and the abundance of Wheeler "signs" of homosexuality.

One conclusion from this study is that research with the Rorschach test that limits itself to objective scores (such as, location, determinants, etc.) fails to capitalize on aspects of Rorschach performance most meaningful in the study of personality.

References

1. Davids, A., M. Joelson and C. McArthur. Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics and normal males. *J. abn. soc. Psychol.*, 1956, *53*, 161-172.
2. Felzer, S. A statistical study of sex differences on the Rorschach. *J. proj. Tech.*, 1955, *19*, 382-386.
3. Knopf, I. J. Rorschach summary scores in differential diagnosis. *J. consult. Psychol.*, 1956, *20*, 99-104.
4. Murray, D. C. An investigation of the Rorschach white space response in an extratensive experience balance as a measure of outwardly directed opposition. (To be published, *J. proj. Tech.*)
5. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch. & J. proj. Tech.*, 1949, *13*, 97-126.

UTILIZATION OF O'REILLY'S OBJECTIVE RORSCHACH AS A SCREENING TEST FOR STATE COLONY JOB APPLICANTS

MARY MARTHA MURPHY

State Colony, Woodbine, N. J.

Introduction

O'Reilly's[3] modification of the Rorschach Test was devised to differentiate among normals, neurotics and psychotics with emphasis upon speed and objectivity of scoring. A scale was set up with twelve possible choices for each card. Four typically psychotic, four typically neurotic and four typically normal responses were selected for each card and the experimental subjects were requested to choose two answers for each inkblot. Scores were assigned to each type of response and cut-off points were established for the purpose of differentiating between (a) normals and neurotics, and (b) psychotics and neurotics. O'Reilly's experimental group of 50 normals was composed of nurses, stenographers, clerks and administrative staff members of the Saskatchewan Hospital, North Battleford, chosen at random from an occupational listing. His neurotic group consisted of 30 recently admitted patients who had been diagnosed psychoneurotics of various stypes by a staff conference composed of medical superintendent, clinical director and nine staff psychiatrists. His psychotic group consisted of 80 patients who had been diagnosed as psychotic by staff conferences.

The present study reports on results of experimentation with O'Reilly's Objective Rorschach as a screening test for job applicants at the State Colony at Woodbine, N. J. Previous investigations have disclosed little evidence of the practicability of other objectively scored and speedily administered modifications of the Rorschach Test. Smith and George[4] found that correlations between the scoring categories of the Harrower Multiple Choice Rorschach and the individual Rorschach were very low. Cronbach[2] reported that the Group Rorschach, objectively scored, failed to predict scholastic success of 200 college students and gave only small correlations with other criteria of the social and emotional adjustment of the sample group. Aldrich and Coffin[1] reported that Harrower-Erickson Multiple-Choice Rorschach ratings of 212 enlistees in the Navy, Marine Corps and Coast Guard failed to disclose it as a practicable method for detecting potentially psychotic individuals.

Method

The Objective Rorschach was individually administered to 60 applicants for employment at Woodbine Colony between March 1, 1956 and September 1, 1956. Investigation of records on March 1, 1957 disclosed that 30 of the original group of job applicants had remained at the Colony as satisfactory employees. Objective Rorschach scores of the original 60 job applicants and the 30 remaining satisfactory employees were compared. Scores of the 30 employees were compared with scores obtained by O'Reilly's normal and neurotic groups and the statistical significance of the differences in scores was ascertained.

Results

The original group of 60 job applicants was composed of 33 males and 27 females. Fourteen of the 30 satisfactory employees were males and 16 were females. The average age of both job applicants and satisfactory employees was 38 with a range of 18 to 57. The average educational level of both groups was 10 grades with a range of 5 to 17 years of formal educational background. Twenty-three of the 30 satisfactory employees were cottage attendants and 7 were professional and clerical workers. Twenty-six (87%) were white and four (13%) were Negro. The mean weighted Wechsler-Bellevue I Vocabulary Scale score of both groups was 10 which is the average verbal subtest score for Wechsler's 20-29 year age groups[5]. Weighted

Vocabulary Scores of the 30 satisfactory employees ranged from 8 to 17 which suggested that verbal intelligence of the group was distributed between low average and superior levels.

Comparison of Objective Rorschach scores of the 60 job applicants with those of the remaining 30 satisfactory employees disclosed little difference in distribution and no statistically significant difference in means. Table 1 indicates that the mean

TABLE 1. COMPARISON OF OBJECTIVE RORSCHACH SCORES OF STATE COLONY EMPLOYEE GROUP WITH SCORES OF NORTH BATTLEFORD NORMAL ANF NEUROTIC GROUPS

Scores	North Battleford normals (n = 50)			Woodbine employees (n = 30)			North Battleford neurotics (n = 30)		
	f	%	Cum. %	f	%	Cum. %	f	%	Cum. %
24	2	4	4						
25	2	4	8						
26	1	2	10	4	13.3	13.3			
27	6	12	22	1	3.3	16.6			
28	7	14	36	1	3.3	19.9			
29	7	14	50	3	10.0	29.9			
30	10	20	70	2	6.6	36.5			
31	4	8	78	2	6.6	43.1	3	10.0	10.0
32	3	6	84				4	13.3	23.3
33	7	14	98	5	16.6	59.7	5	16.6	39.9
34				1	3.3	63.0	2	6.6	46.5
35				4	13.3	76.3	4	13.3	59.8
36	1	2	100	1	3.3	79.6	6	20.0	79.8
37				1	3.3	82.9	5	16.6	96.4
38				2	6.6	89.5	1	3.3	99.7
39									
40									
41				1	3.3	92.8			
42									
43				1	3.3	96.1			
44				1	3.3	99.4			
Mean score	29.4*			32.8			34.4**		
S. D.	2.56			4.85			2.1		

*Difference significant at .0009 level **Difference significant at .10 level

score (32.8) of the Woodbine employee group was almost identical with O'Reilly's cut-off point for normals (33). There was a greater degree of scatter of scores around the mean for the Woodbine employees than for O'Reilly's normal and neurotic groups. Scores of 23.2 per cent of the Woodbine employees ranked in O'Reilly's neurotic category (34 to 37) and 16.5 per cent fell within O'Reilly's psychotic limits (over 37). The difference between mean scores of the Woodbine group and O'Reilly's normals was significant above the .01 level. The mean score of the Woodbine group was not significantly different from O'Reilly's neurotic mean at the 5 per cent level. These results suggest that the Objective Rorschach is impractical as a screening technique for unselected institutional job applicants and appear to support conclusions reported in previous investigations as to the practicability of other objectively scored and speedily administered modifications of the Rorschach Test as a screening technique.

SUMMARY

O'Reilly's Objective Rorschach was administered experimentally as a screening test to 60 job applicants at a state colony. Investigation of records six months later disclosed that 30 of the 60 job applicants had remained as satisfactory employees. Comparison of the employee scores with O'Reilly's norms revealed a statistically significant difference between mean scores of the Colony employees and O'Reilly's normals. The mean score of the Colony group was not significantly different from

O'Reilly's mean for neurotic subjects. Scores of 39.7 per cent of the Colony employees ranked in O'Reilly's neurotic and psychotic categories. This was interpreted as suggesting that the Objective Rorschach is impractical as a screening technique for unselected institutional job applicants.

References

1. Aldrich, C. K. and Coffin, Mable. Clinical studies of psychoses in the Navy. I. Prediction values of social histories and the Harrower-Erickson Test. *J. nerv. ment. Dis.*, 1948, 108, 36-44.
2. Cronbach, L. J. Studies of the Group Rorschach in relation to success in the college of the University of Chicago. *J. educ. Psychol.*, 1950, 41, 65-82.
3. O Reilly, P. O. The Objective Rorschach: a suggested modification of Rorschach technique. *J. clin. Psychol.*, 1956, 12, 27-31.
4. Smith, S. and George, C. E. The Harrower Multiple-Choice Rorschach: a critique. *J. proj. Tech.*, 1954, 18, 507-509.
5. Wechsler, D. *Measurement of adult intelligence.* Baltimore: Williams and Wilkins, 1944.

MOVEMENT IN THE FANTASY OF BRAIN-INJURED (CEREBRAL PALSY) CHILDREN

T. W. RICHARDS*

Louisiana State University

Problem and Procedure

The purpose of this study is to determine whether children brain injured at birth differ from control children in their perception and projection of movement in response to a visual stimulus.

Thirty-two children brain-injured at birth, members of a cerebral palsy group studied in New Orleans, were matched for age, sex and IQ with control subjects. The age range was from 4.5 to 19.5 years; mean IQ for the brain injured was 97.0, for the controls 97.6. Details of the matching and other factors of selection are given in a previous report [2].

The stimulus used was the Projective Movement Sequence (PMS) developed by Lundin [1]. This consists of a motion picture of eight separate scenes of movement of iron filings over a plane surface. Following each scene the subject is asked "what happened?" Reactions to the movement sequences were scored for (a) source of energy (external, internal, disintegrative, static), (b) direction of movement (extensor, flexor, diverse), and (c) content (human, animal, botanical, physiological, nature, etc.). Reliability of scoring of the PMS was determined by duplicate scoring by two observers of twelve of the sixty-four records. They agreed perfectly on 83% of one hundred and twenty responses.

Results

In Table 1, are shown means, medians and sigmas for each of the groups of brain-injured and control children, together with probability data for the significance of differences.

*The author is indebted to Sara W. Hooper, M.A., for assistance in testing, and to Dr. Frank J. Loeffler for help in scoring and statistical analysis.

TABLE 1. MEANS, MEDIANS AND SIGMA FOR VARIOUS SCORING CATEGORIES OF PMS FOR BRAIN-INJURED AND CONTROL SUBJECTS, TOGETHER WITH DEGREE OF SIGNIFICANCE OF DIFFERENCES BETWEEN THEM.

	Brain-injured			Control			Probability	
Factors	Mean	Median	Sigma	Mean	Median	Sigma	t	X^2
Source of Energy								
Internal	5.6	6	3.1	4.1	4	2.8	.05	.10–.20
External	0.7	1	0.8	1.3	1	1.2	.02–.05	.10–.20
Disintegration	1.1	1	1.1	1.0	1	1.1	n.s.	n.s.
Static	2.3	1–2	2.3	3.3	3	2.8	n.s.	.10–.20
Direction								
Extensor	3.1	3	2.4	2.1	2	2.2	.10–.20	.02–.05
Flexor	1.8	2	1.7	2.5	2–3	1.7	n.s.	n.s.
Diverse	1.7	2	2.0	1.4	1	1.6	n.s.	n.s.
Content								
Human	1.0	0–1	1.4	1.3	1	1.5	n.s.	n.s.
Animal	4.1	3–4	2.7	3.9	3–4	2.5	n.s.	n.s.
Inanimate	1.4	0	2.1	0.5	0	1.2	.05–.10	.05–.10
Object	1.1	1	1.5	1.2	1	1.1	n.s.	n.s.
Botanical	1.3	1	1.9	1.2	1	1.2	n.s.	n.s.
Nature	0.5	0	0.8	0.4	0	0.8	n.s.	n.s.
Physiological	0.0	0	0.0	0.7	0	1.3	$<.005$	$<.01$

Comparison of the PMS scores with Rorschach scores for the same children showed that whereas all subjects gave some PMS movement, twenty-eight percent of all subjects gave no movement with the Rorschach. Correlation of movement given by PMS and Rorschach was .25 for brain-injured and .30 for control subjects.

CONCLUSION

A general finding of this study of differences between children brain-injured at birth and controls is that with a motion-picture stimulus of movement the brain-injured children were as spontaneous as controls.

Like the uninjured child the brain-injured subject tends to identify motion as that of animals, humans, plants and objects. He differs from the uninjured in showing total avoidance of movement as a physiological phenomenon, and in using movement that is self-initiated and extensor rather than flexor and initiated by an outside source. The brain-injured child, in his fantasy of activity, is much like the uninjured child, but he shows a significant tendency to disavow the physiological nature of activity, and to experience it as outgoing and active rather than passive. In his fantasy he denies that movement is a matter of his own physiology; rather, movement is an aggressive experience. This substantiates a theory that fantasy serves as wish fulfillment, or at least, compensation for motor handicap.

REFERENCES

1. LUNDIN, WILLIAM H. Projective movement sequences in the study of personality. *J. Proj. Techn.*, 1954, *18*, 208-220.
2. RICHARDS, T. W. and SARA HOOPER. Brain-injury at birth (Cerebral Palsy) and perceptual responses during childhood and adolescence. *J. Nerv. Ment. Dis.* 1956, *123*, 117-124.

DIFFERENTIATING ALCOHOLICS FROM NORMALS AND ABNORMALS WITH THE MMPI[1]

DONALD P. HOYT AND GORDON M. SEDLACEK
Kansas State College *Mental Health Center, Cedar Rapids, Iowa*

PROBLEM

A number of studies purporting either to describe the personality dynamics of alcoholics or to predict chronic alcoholism as a behavioral characteristic have appeared in the literature. (i. e., 3, 6, 7, 8) In general these studies have been seriously limited by utilizing small, non-random samples, by studying only members of Alcoholics Anonymous (probably an unrepresentative group of alcoholics), and by neglecting to differentiate alcoholics from other clinical groups. The present study has as its purpose the identification of personality characteristics of alcoholics which distinguish them from non-alcoholics *and* from other clinical groups.

SAMPLES

Various samples were used in different parts of this study. These samples are listed below, together with an abbreviation code used in the remainder of this report.

A. *Alcoholic subjects.* Two separate samples were used. Both met the following criteria: (a) white males; (b) fifth grade education or its equivalent; (c) diagnosis of chronic alcoholism by a physician and/or a clinical psychologist.

1. The first sample (A-1) consisted of the first 98 cases at the Mental Health Institute in Independence, Iowa,[2] who met the criteria described above.
2. Sample 2 (A-2) was composed of 79 cases from the Willmar State Hospital, Willmar, Minnesota,[3] who met these criteria.

B. *Normal subjects.* Three samples of subjects with no known personality disturbance were used in various phases of this study.

1. The first sample (N–1) was a group of 50 male V. A. on-the-farm trainees, utilized in an unpublished master's thesis by Marsh (8).
2. Sample 2 (N-2) consisted of the 258 normal Minnesota males selected by Hathaway for research on profile patterns of the MMPI (5).
3. The final normal sample (N–3) consisted of the 139 normal Minnesota males and 54 V. A. hospital males used by Hathaway and McKinley in the development of their original scales (9).

C. *Clinical subjects.* Four samples of subjects who had been diagnosed in various clinical groups were utilized.

1. Sample 1 (C-1) was composed of 33 consecutive numbered cases from the Mental Hygiene Clinic at Fort Snelling, Minnesota.[4] In each case the primary diagnostic impression was psychoneurosis.
2. The second sample (C-2) was composed of the 24 male "moderate psychoneurotics" described by Gough (1).
3. Sample 3 (C-3) consisted of Gough's 22 male psychotic patients (1).
4. The final clinical sample (C-4) was Hathaway's group of 710 males with various psychiatric diagnoses selected for research on MMPI profiles (5).

PROCEDURE

Three methods of accomplishing the major objective of the study were employed, all involving the use of the MMPI records of the various subjects. These methods were (a) item analysis, (b) mean profile comparisons, and (c) pattern analysis. Profiles were discarded if the ? or L score was as high as 70 or if the F-K raw score was as large as 9 (2).

Item Analysis. In this method, the percentage of "plus responses" for each of the 566 MMPI items was determined for the A-1 sample. These percentages were compared with those for the N–3 sample (9). Zubin's nomographs (11) were used for determining the significance of the difference. In order to maximize *practical* as well as statistical significance, two criteria for item selection were defined. First of all, only

---

[1]This report is based on the M.S. thesis of the junior author written under the supervision of the senior author.

[2]The cooperation of Dr. Donald Kyer in making these records available is gratefully acknowledged.

[3]These records were made available by Mr. Warren D. Roberts of the Willmar State Hospital, Willmar, Minnesota.

[4]Dr. Carl Jenness and Dr. Daniel Wiener generously provided these records.

those items where the percentage difference was significant beyond the .01 level were included. Secondly, of this group, only those where the percentage difference reached or exceeded 15 per cent were selected.

Mean Profiles. Samples N–1, C–2, and C–3 were compared with the combined A–1 and A–2 samples on the basis of mean scores obtained on each of the nine clinical scales of the MMPI.

Pattern Analysis. The frequency with which various MMPI patterns occurred in the N–2, C–4, and combined A–1 and A–2 samples were compared.

Results and Discussion

Item Analysis. Using the criteria described above, a total of 68 items were obtained. These items are listed in Table 1, along with the scoring direction. Responses of the

Table 1. Items Included in the Alcoholic Scale, Booklet Number, and Keyed Direction

Booklet Number	Item
26	I feel that it is certainly best to keep my mouth shut when I'm in trouble. (F)*
39	At times I feel like smashing things. (F)
46	My judgment is better than it ever was. (F)
61	I have not lived the right kind of life. (T)
94	I do many things which I regret afterwards (I regret more or more often than others seem to). (T)
95	I go to church most every week. (F)
100	I have met problems so full of possibilities that I have been unable to make up my mind about them. (T)
102	My hardest battles are with myself. (T)
127	I know who is responsible for most of my troubles. (T)
131	I do not worry about catching diseases. (T)
140	I like to cook. (T)
144	I would like to be a soldier. (F)
145	At times I feel like picking a fist fight with someone. (F)
155	I am neither gaining nor losing weight. (F)
215	I have used alcohol excessively. (T)
219	I think I would like the work of a contractor. (T)
222	It is not hard for me to ask help from my friends even though I cannot return the favor. (T)
237	My relatives are nearly all in sympathy with me. (F)
239	I have been disappointed in love. (T)
264	I am entirely self-confident. (F)
287	I have very few fears compared to my friends. (F)
289	I am always disgusted with the law when a criminal is freed through the arguments of a good lawyer. (F)
292	I am not likely to speak to people until they speak to me. (F)
294	I have never been in trouble with the law. (F)
300	There never was a time in my life when I liked to play with dolls. (F)
322	I worry over money and business. (F)
327	My father or mother often made me obey when I thought it was unreasonable. (F)
337	I feel anxiety about something or someone almost all the time. (F)
343	I usually have to stop and think before I act even in trifling matters. (F)
346	I have a habit of counting things that are not important such as bulbs on electric signs, and so forth. (F)
348	I tend to be on my guard with people who are somewhat more friendly than I had expected. (F)
351	I get anxious and upset when I have to make a short trip away from home. (F)
359	Sometimes some unimportant thought will run through my mind and bother me for days. (F)
361	I am inclined to take things hard. (F)
365	I feel uneasy indoors. (F)
366	Even when I am with people I feel lonely much of the time. (F)
375	When I am feeling very happy and active someone who is blue or low will spoil it all. (F)
378	I do not like to see women smoke. (F)
383	People often disappoint me. (F)
386	I like to keep people guessing what I'm going to do next. (F)
387	The only miracles I know are simply tricks that other people play on one another. (F)
411	It makes me feel like a failure when I hear of the success of someone I know well. (F)
415	If given the chance I would make a good leader of people. (F)
420	I have had some very unusual religious experiences. (F)
421	One or more members of my family is very nervous. (F)
427	I am embarrassed by dirty stories. (T)
432	I have strong political opinions. (F)
433	I used to have imaginary companions. (F)
436	People usually demand more respect for their own rights than they are willing to allow for others. (F)
437	It is all right to get around the law if you don't actually break it. (T)
446	I enjoy gambling for small stakes. (T)
459	I have one or more bad habits which are so strong that it is no use in fighting against them. (F)
460	I have used alcohol moderately (or not at all). (F)
465	I have several times had a change of heart about my life work. (T)
472	I am fascinated by fire. (F)
473	Whenever possible I avoid being in a crowd. (F)
477	If I were in trouble with several friends who were equally to blame, I would rather take the blame than to give them away. (T)
483	Christ performed miracles such as changing water into wine. (F)
503	It is unusual for me to express strong approval or disapproval of the actions of others. (T)
505	I have had periods when I felt so full of pep that sleep did not seem necessary for days at a time. (F)
513	I think Lincoln was greater than Washington. (F)
516	Some of my family have quick tempers. (F)
524	I am not afraid of picking up a disease or germs from door knobs. (T)
533	I am not bothered by a great deal of belching of gas from my stomach. (T)
554	If I were an artist I would like to draw children. (T)
555	I sometimes feel that I am about to go to pieces. (F)
558	A large number of people are guilty of bad sexual conduct. (F)
560	I am greatly bothered by forgetting where I put things. (F)

* ( ) indicates the scoring direction for the item.

A–2 group were scored on the scale thus obtained in a cross-validation effort. This sample averaged slightly higher (28.4 *vs.* 25.5) than the A–1 sample. Apparently the rigorous selection technique provided an effective safeguard against validity shrinkage.

Two other questions were investigated. First, will the scale differentiate adequately between alcoholics and normals? Second, will the scale differentiate adequately between alcoholics and other clinical groups? To answer the first of these questions, the MMPI responses of the N–1 sample were scored on the new scale. The mean score of the normal group (20.0) was significantly different from that of both the A-1 and A–2 groups (t values were 9.21 and 14.12). By establishing a critical score at 24, 76 percent of the alcoholics and 80 percent of the normals were correctly identified. The second question was answered only partially. MMPI responces of the C–1 sample were scored on the new scale. The mean score obtained (26.4) was not significantly different from that of the alcoholic samples.

Mean Profiles. For purposes of comparing scores obtained on the nine clinical keys of the MMPI, samples A–1 and A–2 were combined into a single sample of 177 alcoholics. This procedure was not entirely justified, since the A-2 groups did score significantly higher than the A–1 group on Pt, Sc, and Ma. The largest of the mean raw score differences was less than 5, however, and this, plus the fact that the general shape of the mean profiles was very similar for both samples (Pd and D above 60 but below 70, other scales above 50 but below 60) was used to justify combining the two samples in order to obtain a larger N.

Figure 1 provides visual comparisons of the mean profile of the alcoholic samples with that of normals (N–1), moderate psychoneurotics (C–2), and psychotics (C–3).[5] Two general conclusions from these figures are obvious. (a) Normals and alcholics do not show markedly different mean profiles on the MMPI. The chief differentiating scales (Pd and D) suggest at best only modest differences in mean scores. (b) On the other hand, each of the two clinical samples obtained mean profiles widely dis-

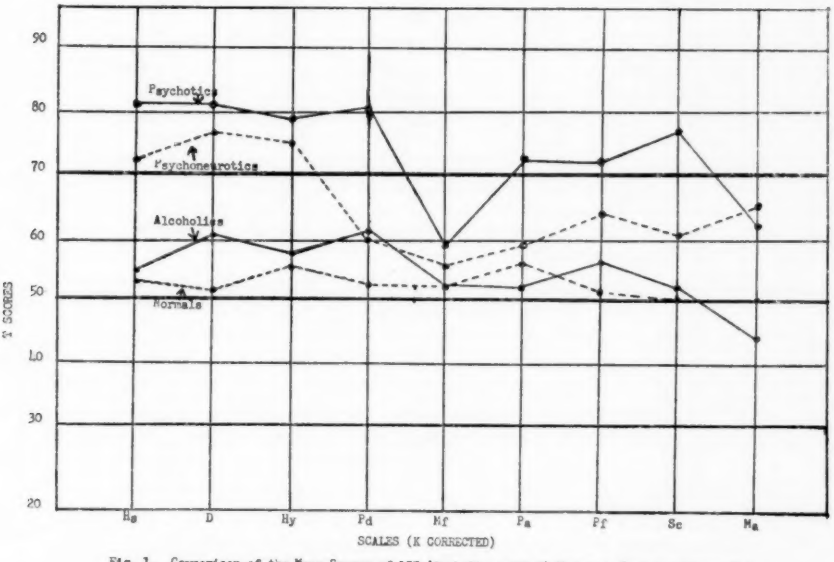

Fig. 1. Comparison of the Mean Scores of 177 Alcoholics with 24 Moderate Psychoneurotics (1), 22 Psychotics (1), and 50 Normals (8).

[5]Reprinted with permission of Dr. Harrison Gough and the *J. clin. Psychol.*

parate with that obtained by the alcoholic group. In the case of the psychoneurotic group (C–2), only Pd and Mf failed to differentiate, and marked differences on the neurotic triad were obtained. The differentiation was even more general in the case of psychotics (C-3), where Mf was the only scale on which similar mean scores were obtained.

Pattern Analysis. An obvious weakness in the preceding analysis is that the mean can conceal more than it reveals. To avoid this pitfall, profiles from the A–1 and

TABLE 2. FREQUENCY OF VARIOUS MMPI PATTERNS** FOR ALCOHOLICS, NORMALS, AND PSYCHIATRIC PATIENTS[6]

Pattern Codes	Alcoholics (N–177) T Scores		Normals (N–258) T Scores		Psychiatric (N–710) T Scores	
	Above 54	Above 70	Above 54	Above 70	Above 54	Above 70
—	0.60	0.00	23.60*	0.00	1.80	0.00
1–	0.00	0.00	1.30*	0.00	0.00	0.00
1–3	0.00	0.00	3.50*	2.30*	6.10*	5.80*
1–4	2.30	0.60	0.40	0.40	1.40*	1.30
1–7	0.60	0.00	0.40	0.40	0.60	0.60*
1–8	0.60	0.00	1.60	0.40	1.30	1.10*
1–9	0.00	0.00	0.80	0.00	0.80*	0.60*
2–1	2.30	2.30	0.80	0.40	6.60*	6.20*
2–3	2.80	1.10	1.20	0.00	3.80	3.20*
2–4	6.60	2.80	0.40*	0.40	2.40*	1.30
2–6	0.60	0.00	0.80	0.40	1.10	1.00*
2–7	3.40	2.30	0.80	0.40	9.90*	9.30*
4–1	4.00	2.30	0.40*	0.00*	0.70*	0.60
4–2	9.60	4.00	0.80*	0.00*	4.10*	3.10
4–3	8.50	1.10	3.50	0.40	2.40*	1.50
4–6	6.20	1.70	0.00*	0.00*	1.50*	1.20
4–7	2.80	1.10	0.40	0.00	1.40*	1.00
4–8	4.00	0.60	1.90	0.80	4.60	3.80*
4–9	14.10	2.30	1.90*	0.80	3.00*	1.50
6–	0.00	0.00	1.20*	0.00	0.00	0.00
6–3	1.10	0.00	0.40	0.00	0.10*	0.00
6–4	0.00	0.00	2.30*	0.00	0.70*	0.30
6–8	0.00	0.00	0.40	0.40	1.00*	1.00
7–	0.00	0.00	1.60*	0.00	0.00	0.00
7–1	0.00	0.00	1.20*	0.00	0.60*	0.60
7–2	1.70	0.60	1.20	0.80	3.80	3.70*
7–8	0.60	0.60	1.20	0.80	2.40*	2.00
7–9	0.00	0.00	1.20*	0.00	0.30	0.10
8–1	0.00	0.00	1.20*	0.00	1.10*	1.00*
8–2	0.00	0.00	0.80	0.00	1.70*	1.50*
8–4	0.00	0.00	1.20*	0.00	1.80*	1.10*
8–6	0.00	0.00	0.80	0.00	1.10*	1.80*
8–7	0.60	0.00	0.40	0.00	3.20*	3.20*
9–	2.80	0.00	9.40*	0.80	1.10	0.30
9–1	0.00	0.00	0.00	0.00	0.60*	0.40
9–2	0.00	0.00	0.40	0.00	1.00*	0.30
9–3	0.00	0.00	0.80	0.00	0.80*	0.10
9–7	0.00	0.00	1.20*	0.40	0.10	0.10
9–8	0.60	0.60	0.80	0.40	2.10*	2.10*

*Significantly different from the alcoholic percentage at the .05 level or lower.

**Patterns where neither the normal nor psychiatric samples differed significantly from the alcoholic have been omitted.

[6]Reprinted with permission of Dr. Starke Hathaway and the University of Minnesota Press.

A–2 samples were coded according to Hathaway's system [4], and the frequency of such codes was compared with Hathaway's data for normal (N–2) and psychiatric (C–4) samples. In accordance with Hathaway's system, special note was taken of T scores above 70. These comparisons are seen in Table 2. Several generalizations are apparent from this table.

1. The alcoholic and the psychiatric groups differed significantly in the patterns exhibited. Using the five percent level of confidence, we would expect 3.25 significant differences by chance; actually, 26 such differences were found.

2. The alcoholics and the normal group differed significantly in the patterns exhibited. Using the five percent level of confidence, we would expect by chance to announce 3.25 significant differences; actually, 18 such differences were found.

3. The percentage of alcoholics exhibiting scores > 70 was significantly less from the psychiatric sample.

4. The percentage of alcoholics exhibiting scores of 70 or more showed no difference from the normal sample. Conclusions 3 and 4 confirmed the earlier findings in terms of the diagnostic significance of profile elevation. Markedly elevated profiles tended to belong to the psychiatric samples much more frequently than to either the alcoholic or normal samples.

5. The Pd scale (scale number 4) was consistently the highest peaked scale for the alcoholic.

The heterogeneity of the psychiatric group in terms of specific diagnoses makes the comparisons in Table 2 less meaningful. It seems reasonable to expect that, in the future, pattern listings for various syndromes will be available, making possible much better comparative information.

Summary and Conclusions

The major purpose of this study was to identify, by use of the MMPI, personality characteristics of alcoholics which differentiate them from normal non-alcoholics and from other clinical groups. Three methods of accomplishing this purpose were employed. First, an item analysis of the MMPI responses of a sample of alcoholics and a sample of normals was performed; the resulting scale was subjected to further analysis. Second, mean scores of alcoholics on the nine clinical scales of the MMPI were compared with mean scores of normals, neurotics, and psychotics. Finally, the frequency of various MMPI patterns for groups of alcoholics, normals, and psychiatric cases were compared. Within the limits of the samples available, it is concluded that:

1. The special scale developed in this study correctly identified 75 to 80 percent of alcoholics and normals under cross-validation conditions.

2. This scale did not discriminate between alcoholics and a group of psychoneurotics.

3. Mean MMPI profiles for alcoholics were, in general, quite similar to that for normals, although Pd and D were somewhat elevated for the alcoholic sample.

4. Samples of psychoneurotics and psychotics obtained much higher mean scores on the MMPI scales than did alcoholics.

5. These four findings suggest that the three groups (normal, alcoholic, and "clinical") can be differentiated by a combination of the MMPI scales and the special "alcoholic" scale.

6. Certain MMPI patterns occurred more frequently in the alcoholic sample than in normal and "psychiatric" samples, particularly in the Pd scale.

7. Lack of clearly defined samples requires caution in interpreting these findings and argues for additional research with better defined groups.

References

1. Gough, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.*, 1946, 2, 23-27.
2. Gough, H. G. The F-K dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, 14, 408-413.
3. Hampton, P. F. The development of a personality questionnaire for drinkers. *Genet. Psychol. Monogr.*, 1953, 48, 55-115.
4. Hathaway, S. R. A coding system for MMPI profile classification. *J. consult. Psychol.*, 1947, 6, 334-337.
5. Hathaway, S. R. and Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis: University of Minnesota Press, 1951.
6. Klebanoff, S. Personality factors in alcoholism as indicated by the Thematic Apperception Test. *J. consult. Psychol.*, 1947, 9, 111-119.
7. Manson, M. P. A psychometric differentiation of alcoholics from nonalcoholics. *Quart. J. Stud. Alcohol.*, 1948, 9, 175-206.
8. Marsh, M. L. *Relationship of aptitude, interest, and personality test scores to performance and job satisfaction.* Unpubl. master's thesis, Kansas State College, 1954.
9. Percentages of "Plus" and "Cannot Say" responses of norm groups to the Minnesota Multiphasic Personality Inventory. Mimeographed paper, University of Minnesota, 1953.
10. Seliger, R. V. Are you an alcoholic? *Alcoholic Hygiene*, 1946, 2, 5-10.
11. Zubin, S. Nomographs for determining the significance of the difference between the frequencies of events in two contrasted series or groups. *J. Amer. Statist. Assoc.*, 1939, 34, 539-544.

THE EFFECT ON SELF DESCRIPTION OF COMBINING ANXIETY AND HOSTILITY ITEMS ON A SINGLE SCALE[1]

MARVIN J. FELDMAN AND SAUL M. SIEGEL

University of Buffalo *Topeka State Hospital*

Problem

In the extensive use of personality questionaires and ratings, little attention has been paid to the possibility that the combination of items being rated may influence ratings. Rather it is generally assumed that ratings depend mainly on the personality of the subject being rated and notions about personality held by the rater. It seems legitimate to inquire, however, whether each item in a test is responded to independently or is to some extent influenced by the nature of the preceding items. The latter view would suggest that the order and composition of items on a scale have some bearing on the personality portrait which emerges from ratings.

Asch (1) has conducted experiments which throw some light on the problem. He found that systematic changes in responding to the same attribute occurred when simple switches were made in the order in which the attribute was presented and when the total composition of the items was altered. For example, "cold" has a quite different meaning for raters when embedded with: intelligent - skillful - industrious - cold - determined - practical - cautious than in the series: intelligent - skillful - sincere - cold - conscientious - helpful - modest. In Asch's study the judges were not rating a real person but rather were forming impressions of a hypothetical case from a list of attributes. With such a task, the interaction of items might exert the maximum influence on rater response. A rating of a real person or a self rating might limit the interaction effect of items as an influence on ratings.

The present study was designed to test the concept that the total composition of a test affects responses to the component parts of the test. The limiting case of self ratings was used. Presumably the rater's self evaluations would be least influenced by side comparisons among items. Hence, if item composition influences self ratings, ratings of others can be expected to be similarly influenced.

---

[1]This article is based upon a paper delivered at the Eastern Psychological Association Meeting, April, 1957. The writers wish to acknowledge the help of Miss Sherry Carpenter and Mr. Richard Remy in computing the statistics.

Materials

To test the effects of item composition on response, three related personality inventories were used: (a) The Taylor Manifest Anxiety Scale (MAS) - 50 items pertaining to anxiety plus 175 buffer items[6]; (b) The Siegel Manifest Hostility Scale (MHS) - 50 items pertaining to hostility plus the same 175 buffer items[4], and (c) Combined form - 50 MAS, 50 MHS, plus 125 of the same buffer items.

Siegel[4] devised the MHS following the same procedure utilized in the construction of the MAS, that is judges agreed that certain MMPI items reflected manifest hostility. In all three forms the scale items were distributed randomly among the buffer items. Anxiety and hostility were chosen as the variables to test our hypothesis because they are clinically relevant variables which might reasonably be expected to exert some mutual influence. Nevertheless, we could not state in advance what direction we expected the effect to take.

Subjects

The subjects were mainly freshmen and sophomores enrolled in an introductory psychology class. The tests were administered as part of the regular class work. Men outnumbered women, 67 to 29, but the respondents were distributed fairly evenly in the subgroups to control any possible effect due to the sex of the subject. The average age of the subjects was 19.8 years with little variability. Some subjects were excluded on a random basis to equalize N's in the subgroups.

Procedure

The tests were administered in two sessions, one week apart. Two separate classes were used for replication purposes. Each class was told that the examiner was collecting normative data on a new personality test. The subjects thought they were all taking the same test. Actually in the first testing, half of each class took the combined form, containing both MAS and MHS items, a quarter of each class took the form with MAS items, and the remaining quarter took the form with MHS items. A week later subjects were tested again ostensibly so that the reliability of the test could be assessed. On the second administration, subjects who previously took the combined form now received one of the other forms containing *either* the MAS or MHS items. Conversely, subjects who took either the MAS or MHS initially now received the combined form. Thus each subject took two of the three forms. All 96 subjects took the combined form, 56 took the separate MHS form and 40 the separate MAS form. None of the subjects suspected that the second test differed somewhat from the first test.

Results

Before presenting the main results, it is of interest to examine the reliability of the measures. It is possible that the reliability of the measures differs in the combined compared with the separate forms of the scales. It is also possible that reliability might be affected by the order in which the forms were taken. With the present design, both test-retest and split-half reliability coefficients could be computed.

Table 1 indicates that the reliability coefficients were fairly high. Test-retest coefficients were somewhat higher than split-half coefficients even when the latter

Table 1. Reliability Coefficients on MAS and MHS

Form	MAS		MHS	
Test-retest:				
separate *vs.* combined form	.86		.78	
separate first *vs.* combined second	.85		.84	
combined first *vs.* separate second	.89		.76	
Split-half:				
separate form	.64	.78*	.66	.80*
combined form	.62	.77*	.69	.82*

*corrected by Spearman-Brown formula

were corrected by the Spearman-Brown formula. Frequently with personality tests the reverse holds true. Neither the form of the scale, nor the order in which tests were taken seems to have affected the reliability. The present test-retest coefficients compare favorably with other results reported[5]. The present split-half correlation on the MAS is somewhat lower than results reported by Hilgard, Jones, and Kaplan[2], but on the MHS, the coefficients are about the same as was found in the initial study[4]. In general, however, the reliability coefficients are high enough to permit confidence in using these measures in the present study.

To test the main hypothesis, several analyses of variance were computed using a Type III, mixed design described by Lindquist[3]. This design conveniently tests the main assumptions of the study while also permitting an analysis of two control factors inherent in the study, namely the study of possible differences between classes—a replication factor—and possible differences stemming from the order in which the two forms of the test were taken, that is the separate MAS or MHS first or the combined form first.

Table 2 presents the results from the analysis of the MAS scores. None of the F ratios were significant. Thus it seems that MAS scores were not affected by inclusion of hostility items. The two classes gave essentially the same results, and the order in which forms were taken had no effect.

A similar analysis was carried out to determine the influence of anxiety items on MHS scores. Table 2 indicates that MHS scores did not vary in terms of the control

TABLE 2. ANALYSIS OF VARIANCE OF MAS AND MHS SCORES

Source	MAS Scores df	MAS Scores ms	MAS Scores F	MHS Scores df	MHS Scores ms	MHS Scores F
Between S:	39			55		
B – order	1	2.12	.02	1	58.58	.95
C – class	1	52.82	.45	1	23.22	.38
BC	1	.10	.00	1	12.22	.02
error (b)	36	116.54		52	61.41	
Within S:	40			56		
A – form of test	1	6.62	.81	1	23.22	3.61*
AB	1	2.10	.25	1	27.01	4.20**
AC	1	6.60	.87	1	9.73	1.51
ABC	1	.13	.01	1	.44	.01
error (w)	36	8.22		52	6.43	
Total	79			111		

*significant $> .05 < .10$ level
**significant at .05 level

variables, different classes or order in taking the two forms of the test. The main hypothesis is born out to some extent, however, for MHS scores were reduced somewhat by inclusion of anxiety items. Thus the analysis of variance comparing the separate with the combined form of the MHS yields an F ratio of 3.61 which is significant between the 5 and 10% level. Furthermore, the effect is more pronounced when subjects gain familiarity with the scale by taking the separate MHS form first, for there is a significant interaction at the .05 level between form of the scale and order in taking the two forms. To facilitate comparisons with other studies using these scales, Table 3 presents the means and standard deviations of the various subgroups.

The results obtained suggest that combining subscales had only a minimal effect on responses to either scale. It seemed possible that the effect might be stronger in the second half of the scale after subjects built up a set on the first half. To test this possibility the scales were all rescored on the last 25 items only. The same statistical operations were then carried out. The results, however, were in substantial agreement with the main findings so do not seem to warrant a detailed analysis. Again

TABLE 3. SUBGROUP MEANS AND STANDARD DEVIATIONS ON THE MAS AND THE MHS

Group	MAS (N = 10 per group)		MHS (N = 14 per group)	
	M	S	M	S
Separate form first – class A	13.7	8.2	15.7	5.8
Combined form second – class A	13.3	7.8	13.4	6.0
Combined form first – class A	13.7	7.4	14.1	6.7
Separate form second – class A	14.1	5.5	13.4	5.0
Separate form first – class B	15.9	6.0	15.0	4.3
Combined form second – class B	14.5	7.1	13.6	5.4
Combined form first – class B	15.9	9.6	11.8	5.7
Separate form second – class B	15.0	7.9	12.6	5.7

the inclusion of hostility items seemed to have little effect on responses to MAS while inclusion of anxiety items lowered scores on the MHS, but the F ratio was still only significant at the .05 to .10 level.

SUMMARY

This study investigated the possibility that responses to test items might be influenced by inclusion of different item content in the same test. It was found that scores on a scale of manifest hostility were lowered when the total test also included anxiety items. The effect was not general for scores on a scale of manifest anxiety remained constant when hostility items were added to the total test.

REFERENCES

1. ASCH, S. E. Forming impressions of personality. *J. abn. soc. Psychol.*, 1946, *41*, 258-290.
2. HILGARD, E. R., JONES, L. V. and KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, *42*, 94-99.
3. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*, New York: Houghton Mifflin Co., 1953.
4. SIEGEL, S. M. The relationship of hostility to authoritarianism. *J. abn. soc. Psychol.*, 1956, *52*, 368-372.
5. TAYLOR, JANET A. A personality scale of manifest anxiety., *J. abn. soc. Psychol.*, 1953, *48*, 285-290.

FACE VALIDITY IN THE D SCALE OF THE MMPI

RAYMOND J. MC CALL[1]

Marquette University

PROBLEM

In an earlier study Doleys (1) found the D scale of the MMPI to be of moderate validity ($p < .05$) in discriminating a group of 41 depressives from a matched group of non-depressive psychotics. An item analysis of his results showed, however, that sizable portions of the D scale were of little or no differential value while other blocs of items seemed quite effective. Aside from noting the inutility of the so-called "correction" items (2), Doleys was unable to group the discriminating and the non-discriminating items into very meaningful classes.

[1]Thanks are due to E. J. Doleys Jr., who compiled the original data, and to C. L. Klingberg and P. J. Vitu for assistance with statistical computations.

Method

Independently of Doleys' analysis, the writer decided to test on these data the hypothesis that the most discriminating items would be those which come closest to describing the known symptomatology of depression; that items less essential to the syndrome would be less discriminating; and that, finally, items inconsistent with or irrelevant to the depressive syndrome would contribute least to distinguishing diagnosed depressives from non-depressive psychotics. Accordingly, the 60 items of the D scale were broken down *a priori* into 3 groups: 26 "face valid" (FV) items, 22 "congruent" (CG) items, and 12 "irrelevant" (IR) items. These items as scored for depression may be identified as follows:

I. "Face Valid" Items

B2	Not understanding what I read as well as formerly
E41	Not enjoying different kinds of play and recreation
F36	Wishing I could be as happy as others
F38	Not feeling that life is worthwhile
F39	Finding daily life not interesting
F42	Having difficulty in starting to do things
F44	Having periods when I could not get going
F45	Brooding a great deal
F51	Crying easily
G7	Not caring what happens to me
G12	Not being happy most of the time
I27	Finding it hard to keep my mind on a task
I37	Feeling useless at times
A2	Not being in as good physical health as friends
A3	Having felt better before
A6	Worrying about health
B12	Not having a good appetite
B18	Troubled by constipation
B30	Not going to sleep without troublesome thoughts
A22	Not being as capable and smart as most others
A24	Judgment *not* better than ever
I39	Lacking self-confidence
G18	Not having periods of unusual cheerfulness
G23	Not feeling full of energy
G25	Not feeling excitedly happy even when things go wrong
C17	Not being as able to work as ever

II. "Congruent" Items

G38	Worrying about locked doors and windows when leave home
I4	*Not* teasing animals
B36	*Not* dreaming frequently about things best kept to myself
D49	Blaming a person for taking advantage of someone
F34	Being terribly hurt by criticism
H43	Worrying about catching diseases
J50	Never laughing at a dirty joke
AI	Not having been well most of the last few years
A4	Either gaining or losing weight
A40	Often feeling weak all over
B28	Sleep fitful and disturbed
B31	Being easily awakened by noise
G31	Never feeling like smashing things
G32	Never feeling like picking a fist fight
J51	Never feeling like swearing
A8	Believing I am more nervous than most
A27	Memory not seeming all right
H55	Being afraid of losing my mind
C48	Not liking to flirt
E24	Passing by other than close friends unless they speak first
E36	Not being a good mixer
I34	Working under a great deal of tension

III. "Irrelevant" Items

*A18	Having had a fit or convulsion
*B3	Not sweating easily on cool days
*B4	Not sweating when embarrassed

*B6	Not having spells of hay fever or asthma
*B8	Never having urinated or coughed up blood
*D10	Not going to church regularly
*D18	Not believing in the second coming of Christ
*D22	Not believing things are turning out as predicted by the Biblical prophets
*G37	Not keeping on at things till people lose their patience
C33	Not agreeing that most people are hard to convince of truth
A15	Being troubled by attacks of nausea and vomiting
C31	Not standing in other people's way because of principle

* "Correction" item

This subclassification of the D scale items may be regarded as purely "intuitive", since no formal attempt at empirical validation was made. The fact that 9 of 12 items classified as "irrelevant" are from the 11 "correction" items of Hathaway and McKinley, or that most of the FV items may be found attributed to depressives in the standard text books, reveals something about the author's procedure but is in no way crucial to the empiriological design.

Results

Considering mean number of items answered in the scored direction by the depressives and the non-depressive psychotics, *t*-ratios set forth in Table 1 show that the 26 FV items differentiated the two groups at least as well as did the scale as a whole.

Table 1. Mean Differences in Number of Scored D Scale Items for Depressives and Non-Depressives

Item Classification	Mean diff.	t
Total D Scale (60 items)	2.27	2.01*
FV items (26 items)	3.11	2.37*
CG items (22 items)	2.18	1.04
FV + CG items (48 items)	2.69	2.12*
IR items (12 items)	.59	.24

*$p < .05$

To test further the apparent trend toward lesser discriminating value as the items departed from face validity (which the *t*-test suggested but did not prove), the differences were analyzed non-parametrically according to the Wilcoxon signed-rank technique. The results are summarized in Table 2.

Table 2. Signed-Rank Discrimination of Depressives From Non-Depressives in the MMPI D Scale

Item Classification	Value of normal deviate	*p*
Total Scale	–3.64	< .01
FV items	–3.50	< .01
CG items	–2.16	< .05
IR items	– .81	> .10

The proportion of depressives and non-depressives answering each type of item in the scored direction was also analyzed by the analysis of variance technique (Table 3), each proportion having been first transformed into an arc sine equivalent. As a further check, the analysis was repeated (Table 4), this time considering proportion of items of each type answered in the scored direction (as opposed to proportion of persons answering each type of item.). The results cohere perfectly, and are consistent with the trend indicated by the signed-rank test.

TABLE 3. ANALYSIS OF VARIANCE FOR PROPORTION OF DEPRESSIVES AND NON-DEPRESSIVES SCORING POSITIVELY ON THE MMPI D SCALE

Source	Sum of Squares	df	Mean Sq.	F
Item Type	.36	2	.18	8.02**
Diagnostic category	.11	1	.11	4.74*
Interaction	.02	2	.01	
Within	2.55	114	.02	
Total	3.04	119		

**$p < .001$
*$p < .05$

TABLE 4. ANALYSIS OF VARIANCE FOR PROPORTION OF ITEMS SCORED POSITIVELY IN THE MMPI D SCALE

Source	Sum of Squares	df	Mean Sq.	F
Item Types	.81	2	.41	13.00**
Diagnostic categories	.17	1	.17	5.49*
Interaction	.09	2	.045	
Within	7.51	240		
Total	8.58	245		

**$p < .001$
*$p < .05$

SUMMARY AND CONCLUSIONS

The 60 items in the depression scale of the MMPI were shown to be differentially effective in distinguishing depressive from a matched group of non-depressive psychotics in proportion to their "face validity", as determined *a priori*.

The implication that the scale might be improved for purposes of differential diagnosis by elimination of "irrelevant" items is probably worth further investigation.

BIBLIOGRAPHY

1. DOLEYS, E. J. JR., The validity of the depression scale of the MMPI. Unpublished master's thesis, De Paul University, 1954.
2. HATHAWAY, S. R. and MCKINLEY, J. C. A multiphasic personality schedule III. The measurement of symptomatic depression. *J. Psychol.*, 1942, *14*, 73-84.

THE TAYLOR MANIFEST ANXIETY SCALE IN DIFFERENTIAL DIAGNOSIS[1]

HAROLD RUBIN AND ALAN H. TOWNSEND

Veterans Administration Hospital, Lebanon, Pennsylvania

University of Michigan

Problem

The Taylor Manifest Anxiety Scale (MAS) was originally devised as a measure for exploring the relationship between performance in learning situations and effective drive level[7]. More recently, however, the MAS has become rather widely used as a measure of clinical anxiety, and numerous studies evaluating its validity have been reviewed by Taylor[7]. In general they seem to suggest that the MAS bears some relationship to clinically observed anxiety.

In one of the few studies on hospitalized psychiatric patients, Buss, *et al.*[2] reported a correlation of .60 between the MAS and over-all ratings of anxiety obtained from a pool of judges. Taylor states[7] ". . . the experimental results obtained with the anxiety scale might also hold for groups selected according to clinical criteria." In a recent study by Siegman[6], the MAS scores of various psychopathological groups indicated that patients with a diagnosis of anxiety reaction had significantly higher MAS scores than patients in any of the other groups, and also that patients with a diagnosis of psychopathic personality had significantly lower scores than any other group. Rubin, *et al.*[5] in a study on hospitalized psychotics found that patients who were rated as well-adjusted on the Hospital Adjustment Scale[4] were relatively free of expressed manifest anxiety symptoms, obtaining low MAS scores. On the other hand, poorly adjusted patients were found to express both little and much manifest anxiety, and their MAS scores were unpredictable, ranging from very low to very high.

It would seem to follow then that patients with psychiatric diagnoses of "anxiety reaction" should obtain higher MAS scores than patients with other psychoneurotic diagnoses but in whom manifest anxiety was not the predominant symptom[1]. It would also appear that better adjusted patients obtain lower MAS scores than the more poorly adjusted ones. The present study attempts to explore these hypotheses in evaluating the use of the MAS in the differential diagnoses of hospitalized patients.

Subjects and Procedures

Subjects were 239 hospitalized, male veterans on the admission ward of the neuropsychiatric service. Eighty-two subjects were diagnosed as *psychoneurotics* and 157 subjects as *schizophrenics*. Of the psychoneurotics, 57 were diagnosed as "anxiety reaction" and 25 carried diagnoses other than "anxiety reaction"[1]. The schizophrenic group consisted of 82 patients who were scheduled to receive electroshock therapy and 75 patients who were admitted at a later date, when shock therapy was no longer a prevalent form of treatment on the neuropsychiatric service.

The shock group was included here and separately, because it was felt that their findings would contribute meaningfully to this study. It seemed that the decision to administer electroshock treatment was based in part on the clinical symptoms presented at the time of admission and these in turn seemed to reflect the type of item sampled by the MAS. All of the diagnoses were made by the psychiatric staff of the hospital following a period of observation on the admission service. For the purpose of simplification, the two groups of neurotics will hereafter be referred to as the *anxiety* and *non-anxiety* groups; the two groups of schizophrenics will be referred to as the *random schizophrenic* and *shock* groups.

---

[1]From the VA Hospital, Lebanon, Pennsylvania.

Results

In Tables 1 and 2 are presented the mean MAS scores and *t*-values for the different psychopathological groups. Among the psychoneurotics, those patients diagnosed as "anxiety reaction" failed to show a significantly higher MAS score than those with a non-anxiety diagnosis, with the former group actually slightly lower. The MAS mean scores were 21.9 for the *anxiety* group and 23.2 for the *non-anxiety* group. Among the psychotics, there was a significant difference between the MAS scores of those patients selected at random and the MAS scores of patients selected for electroshock therapy, the latter group scoring higher. The *random schizophrenic* group obtained a mean MAS score of 14.9 and the *shock* group a mean MAS score of 20.3. When the total psychoneurotic group was compared with the total psychotic group, there was a significant difference between their mean MAS scores, with the former receiving the higher scores. For the psychoneurotics the MAS mean was 22.4 and for the psychotics it was 17.7.

Table 1. Mean Taylor MAS Scores

Group	N	Mean	S.D.
Psychoneurotic	82	22.4	10.78
Anxiety	57	21.9	9.82
Non-anxiety	25	23.2	12.68
Schizophrenic	157	17.7	15.59
Random	75	14.9	9.27
Shock	82	20.3	9.17

Table 2. *t* Values of the Differences Between Mean Taylor MAS Scores for Different Psychopathological Groups

Psychopathological groups	*t*	df
Psychoneurotic vs. schizophrenic	3.09**	235
Anxiety vs. non-anxiety	.496	80
Shock vs. random	3.67**	155
Anxiety vs. random	4.17**	130
Non-anxiety vs. random	3.45**	98
Anxiety vs. shock	.774	137
Non-anxiety vs. shock	1.25	105

**signif. at .01 level

Further inter-comparisons of the four groups indicate that significant differences occur only when the random schizophrenic group is included. By themselves, the three other groups (anxiety, non-anxiety and shock) fail to obtain significantly different MAS scores, as can be observed from Table 1.

Discussion

That the MAS as a measure of "anxiety" is inconsistent with psychiatric diagnoses is not surprising, in view of the different frames of reference on which they are based. The diagnosis of "anxiety reaction" usually follows the APA diagnostic manual[1] which claims this condition to be ". . . characterized by anxious expectation and frequently associated with somatic symptomatology." On the other hand, the MAS measures symptoms of a more behavioral nature (restlessness, irritability, etc.) and it is dependent on "the subject's ability and willingness to express such symptoms or complaints".[5]

In a recent study on anxiety in clinical situations, Buss, *et al.*[2] reported a much lower correlation between the MAS and observed (by rater) categories of manifest anxiety than between the MAS and reported (by subject) categories of manifest anxiety. It was also reported that the highest relationships were found between the

MAS and physiological concomitants of manifest anxiety as reported by the subject. Siegman[6] attributes a low correlation between the MAS and a manifest anxiety rating scale to the fact that the MAS is "loaded with chronic somatic manifestations of anxiety (headaches, stomach troubles, etc.)" which in an item analysis failed to differentiate between his psychopathological groups and which he concludes is thus "an undesired source of variance".

Summary

The present study attempted to evaluate the role of the Taylor Manifest Anxiety Scale in differential psychiatric diagnosis. Subjects were 239 hospitalized psychiatric patients of whom 82 were diagnosed as psychoneurotic and 157 as schizophrenic. Results seemed to indicate that although the MAS made some contribution to the psychiatric evaluation of these patients, it was limited in its ability to discriminate between patients carrying different diagnostic "labels", including *anxiety reaction*.

References

1. American Psychiatric Association. *Diagnostic and statistical manual.* Washington, D. C.: American Psychiatric Association, 1952.
2. Buss, A. H., Wiener, M., Durkee, A., and Baer, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, 19, 125-129.
3. Kelly, G. A. *The psychology of personal constructs.* Vol. 1. *A theory of personality.* New York: Norton, 1955.
4. McReynolds, P. and Ferguson, J. T. *Clinical manual for the Hospital Adjustment Scale.* Stanford, Calif.: Stanford U. Press, 1953.
5. Rubin, H., Schneiderman, L., Hallow, W. C., and Jones, R. J. Manifest anxiety in psychotics. *J. clin. Psychol.*, 1956, 12, 94-96.
6. Siegman, A. W. Cognitive, affective and psychopathological correlates of the Taylor Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, 20, 137-141.
7. Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, 53, 303-320.

PREDICTION OF REHOSPITALIZATION USING THE MMPI[1]

PETER F. BRIGGS

University of Minnesota

Peterson[2] has offered a six-point scale which bears on the diagnosis of subclinical schizophrenia and the prediction of future hospitalization derived from MMPI profiles. The present report describes a study of the power of this scale to predict psychiatric rehospitalization.

Procedure

Five diagnostic samples were drawn from the inpatient records of the psychopathic hospital at the University of Minnesota. Four of these samples have been presented by Schofield and Briggs[3] in a study of methods of clinical prediction. The diagnostic groups are manic-depressive (17), involutional depression (17), schizophrenia (30), psychoneurosis (35), and mixed psychoneurosis (42). The requirement for selection in the first four groups was the presence of both intake and discharge Multiphasic profiles. The fifth group has been described by Holt[1]. It included all patients having a diagnosis of mixed psychoneurosis admitted to the hospital between 1942 and 1948 and having intake MMPIs.

[1]Partly supported by a grant-in-aid of research from the University of Minnesota Graduate School.

Patients in all five samples were cleared through the state's Central Index which records all cases of hospitalization in state institutions. For the first four groups the median follow-up period was five years and ten months. At that time, 30% of these samples had been rehospitalized. Among the mixed psychoneurotics, 31 per cent had been rehospitalized within a median follow-up period of eleven years.

All admission MMPI profiles were scored on the six points suggested by Peterson: (a) T scores of four or more of the clinical scales over 70, (b) F > 65, (c) Sc > Pt, (d) Pa or Ma > 70, (e) Pa or Sc or Ma > Hs and D and Hy, (f) D > Hs and Hy.

Results

Inspection of the results suggests that rehospitalization becomes more likely as the number of Peterson scale points increases. For the groups combined, this trend is not reliable because the schizophrenia group appears to respond with a reverse trend at the highest score. Table 1 shows the number of cases rehospitalized in each

Table 1. Number of Cases Rehospitalized Dichotomized Between Three and Four on Peterson's Scale for Five Diagnostic Groups*

Factors	Maniac Depressive	Schizophrenic	Involutional Depression	Psychoneurosis	Psychoneurosis Mixed
Number	17	30	17	35	42
No. Rehospitalized	5	12	5	8	13
% < 4 Rehosp.	27	50	27	13	24
% ≥ 4 Rehosp.	33	31	50	45	63

*One case has been removed from this table because of its double counting, reducing the N from 101 in the first study [3] to 99 in this study.

of the five groups for scale scores dichotomized between score 3 and 4. The table shows clearly that the likelihood of rehospitalization among neurotics is greater for those with more than three scale points. Taking the same point of dichotomy for the Peterson scale, the combined neurotic groups have a rate of 19 per cent rehospitalized for those with scores below 4 as opposed to a rate of 52 per cent rehospitalized for those with scale scores of 4 or more ($X^2 = P < .01$).

One further comment seems deserved. Among the schizophrenics with four or five points, four out of ten patients were rehospitalized; among those schizophrenics with six points, only one of six cases was rehospitalized. In this group, bizarreness of the symptom picture seems to be associated with good outcome. Unfortunately, the diagnoses on rehospitalization are not known. Such knowledge would provide a better test of Peterson's work.

Summary

A follow-up study was done on five diagnostic samples of clinical cases to determine whether Peterson's scale utilizing MMPI profiles can be used to predict the incidence of rehospitalization. The results, utilizing median follow-up periods ranging from five years ten months to eleven years, indicate reliable predictive utility for the Peterson scale.

References

1. Holt, Shirley M. Prognostic Factors in Psychoneurosis Mixed Type. Unpublished doctoral dissertation. University of Minnesota, 1955.
2. Peterson, D. R. The diagnosis of subclinical schizophrenia. *J. consult. Psychol.*, 1954, 18, 198-200.
3. Schofield, W. and Briggs, P. F. Criteria of therapeutic response in hospitalized psychiatric patients (in preparation).

NOTE ON A RAPID SCORING PROCEDURE FOR THE CARD FORM MMPI

GEORGE T. WILCOX

Adult Psychiatric Clinic, Dayton, Ohio

A system for encoding significant response and scale on each MMPI card to permit scoring without tabulating has recently come to the writer's attention. The bottom edges of the clinical cards are marked in such a way that when all the cards are inverted in the box and aligned against one side, the marks will form a series of columns, each aligned with a scale designation on the index cards. Significant response is indicated by color, and scale by position, of the marks. The raw scores for each clinical and validation scale plus the raw scores for one or more supplementary scales can then be counted up directly from the cards.

The bottom portion of the front surfaces of the index cards marked "False" and "Cannot Say" are marked off into 12, 13 or 14 spaces, leaving ½" at each end unmarked because of the corner notches on the clinical cards. The spaces are then labelled with the conventional abbreviations of the clinical and validation scales, plus one or two supplementary scales if desired. This labelling could be done in red ink on the "False" card and in blue ink on the "Cannot Say" card, to avoid confusion in scoring.

Using an analysis of the items in the scales, such as[1], an ink mark about 1/16" long is made on the bottom edge of each card listed for a particular scale, the mark aligned with the space allotted to that scale on the index cards. When the significant response is "true", the mark is made in red ink. When it is "false", blue or black ink is used. On the 5 Mf scale items differentially scored for testee's sex, we made 3 closely spaced dots, the color of the center dot indicating the significant response for females, and the two outer dots' color indicating the significant response for males.

To score the record, one simply counts the number of blue marks on the cards in front of the "Cannot Say" card, and the number of red marks in front of the "False" card, making a cumulative total. This is done for each column. The totals are transferred directly to the bottom of the profile chart.

We have found that, with care in marking, and with no more than 13 or 14 scales used, there is enough separation between columns of marks to minimize confusion while scoring. The marking is a laborious process which can be expected to take several days' time. It is best performed by a person with considerable clerical aptitude. It is advisable to check the marking by scoring one record by both this and the conventional method. We have used this system for about a year. So far no patient has indicated awareness of the marks on the cards.

Among the advantages of this system, the rapid scoring is outstanding. By this system, the card form MMPI can be scored more quickly than the group form. This, of course, frees the cards promptly for the next administration. Another advantage is that the responses that produced a deviant score on the profile can be picked out directly from the deck of cards for individual consideration. The major disadvantage is that there is no permanent itemized record of the testee's responses. The only other disadvantage which has come to light so far is that one is limited to scoring for the scales that are coded. The seriousness of these drawbacks would vary with the situation in which the test is used.

REFERENCE

1. Unpublished mimeo. #CP-9C, from Dept. of Neuropsychiatry, Clin. Psy. Sec., M.F.S.S., B.A.M.C., Ft. Sam Houston, Texas.

THE CLINICAL APPLICATION OF SEQUENTIAL ANALYSIS

ARNOLD TREHUB AND ISIDOR W. SCHERER

Veterans Administration Hospital, Northampton, Mass.

Introduction

Sequential analysis is a relatively recent statistical development [1, 2, 3, 4], and appears of particular interest in the clinical setting. In the clinic or hospital, information of considerable potential value is recorded each day and filed away with the hope that some day it will be disinterred and subjected to statistical assessment. When the number of cases has grown and it is decided to analyze the data, the task of searching the records, transcribing the pertinent information, and making the statistical analysis may be a formidable one. In sequential analysis, however, we have available an elegant technique which allows us to assess information in terms of statistical hypotheses as each case is recorded. One of the potentially most useful applications of sequential analysis to clinical problems is in the testing of proportions. It is this application which is presented here. This paper has been limited to the goal of indicating some of the possible clinical uses of sequential analysis and some of the advantages inherent in this technique.[1]

In setting up a plan of sequential analysis, alternate hypotheses must first be explicitly stated in terms of proportions. Conventional symbols for the alternate hypotheses are H_0 and H_1. If we are interested, for example, in the ratio of human movement responses to color responses given by hysterics on the Rorschach, we might set up alternate hypotheses as follows. Sum C > M occurs in 70% of the cases. Then $H_0 = .70$. Sum C > M occurs in 80% of the cases. Then $H_1 = .80$. After stating H_0 and H_1, we must decide what proportion of the time we are willing to reject H_0 when in fact it is true, and what proportion of the time we are willing to accept H_0 when in fact H_1 is true. The statements of these two proportions are called a values and β values respectively. Having decided the values of H_0, H_1, a, and β, we may then draw on any standard graph paper two parallel decision lines, the slope, separation, and course of which are determined by solving at least two values for y at arbitrary x points in each of the following formulae:

$$y \log_e \frac{P_1}{P_0} + x \log_e \frac{1 - P_1}{1 - P_0} = \log_e \frac{1 - \beta}{a} \qquad (1)$$

and

$$y \log_e \frac{P_1}{P_0} + x \log_e \frac{1 - P_1}{1 - P_0} = \log_e \frac{\beta}{1 - a} \qquad (2)$$

where P_1 is the proportion in hypothesis H_1, and P_0 is the proportion in hypothesis H_0; y refers to points on the ordinate of the graph, while x refers to points on the abscissa. The line through coordinates computed from formula (1) will be called the "accept H_1" line, and the line through coordinates computed from formula (2) will be called the "accept H_0" line. Table 1 gives xy coordinates which can be used directly to draw H_1 and H_0 lines at 10 percent intervals, and at the risks of $a = .05$, $\beta = .05$.

---

[1]For a more general and theoretical presentation of sequential analysis and related concepts such as operating characteristics and average sample number, the reader is referred to the publications cited above, in particular the Columbia University Group publication [1], and Wald [4].

TABLE 1. VALUES OF y AT ARBITRARY VALUES OF x FOR ALTERNATE HYPOTHESES ($a = .05, \beta = .05$).

Alternate Hypotheses		Arbitrary Values of x				
$H_1 : P_1$	$H_0 : P_0$	0	5	10	20	40
.20	.10	4.2 −4.2			7.6 −.8	11.0 2.6
.30	.20	7.3 −7.3			13.9 −.7	20.5 6.0
.40	.30	10.2 −10.2			20.9 .5	31.6 11.2
.50	.40	13.2 −13.2		21.4 −5.0	29.5 3.1	
.60	.50	16.2 −16.2		28.4 −3.9	40.7 8.3	
.70	.60	19.1 −19.1	28.5 −9.8	37.8 .4		
.80	.70	22.0 −22.0	37.1 −6.9	52.2 8.2		
.90	.80	24.9 −24.9	34.4 4.5	83.8 33.9		

HYPOTHETICAL APPLICATIONS

Following are some examples of the possible application of sequential analysis to typical clinical problems.

1. It is known that Treatment A results in a clinical improvement rate of 60 percent for patients diagnosed as paranoid schizophrenics. A new treatment B has been suggested. It has been decided that if the hypothesis that treatment B is 60 percent effective can be rejected in favor of the alternate hypothesis that treatment B is 70 percent effective at an acceptable level of confidence, B will be adopted as the treatment of choice. In this case, the following sequential analysis plan might be set up.

   a. Let H_1 and H_0 refer to the alternate hypotheses of improvement rate with treatment B. Thus, $H_1 : P_1 = .70$, and $H_0 : P_0 = .60$.

   b. Let risk $a = .05$, and risk $\beta = .05$.

On the basis of values chosen for H_1, H_0, a, and β, two decision lines are drawn (see Table 1, Fig. 1). Each patient diagnosed as paranoid schizophrenic who completes treatment B is classified as improved or unimproved. Let us assign the number 1 to an improved case, and 0 to an unimproved case. For each 1 recorded, a line is drawn one unit in the vertical direction within the channel formed by the "accept H_1" line and the "accept H_0" line. For each 0 recorded, a line is drawn one unit in the horizontal direction. Each successive line starts at the point reached by the preceding entry. As long as the entries remain within the channel, classification and plotting of cases continues. When either decision line is crossed by an entry, sampling stops and the hypothesis represented by the line crossed is accepted. Now if the data come in and are classified in the following sequence: 11101111110111001111111 11011111111, we can see from Fig. 1 that H_1 (70 percent improvement) is accepted as more likely ($a = .05$) than H_0 (60 percent improvement). In this instance, treatment B would be adopted as the treatment of choice.

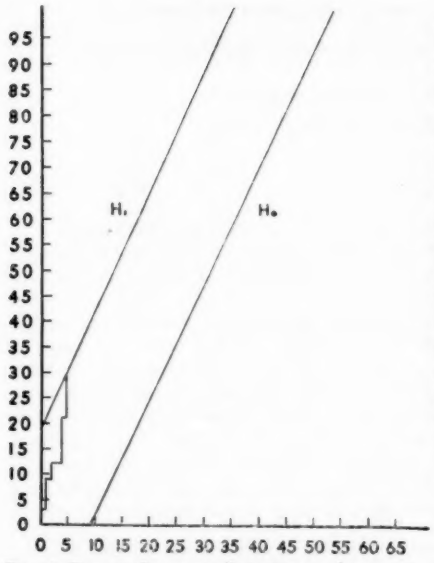

FIG. 1. SINGLE CHANNEL SEQUENTIAL ANALYSIS. DECISION: ACCEPT $H_1 : P_1 = .70$.

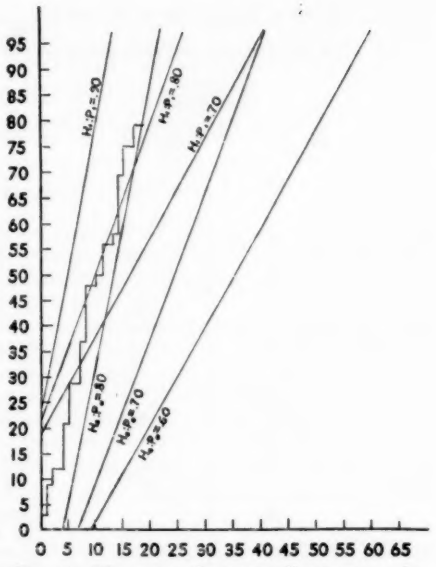

FIG. 2. MULTIPLE CHANNEL SEQUENTIAL ANALYSIS. DECISION: ACCEPT $H_1 : P_1 = .80$ AND $H_0 : P_0 = .70$.

While in this example H_0 is based upon the proportion of improvement resulting from an established treatment, in the general case H_0 could, of course, be set by an obtained proportion in any relevant control group.

2. Suppose our problem is not one of choosing between two treatments, but of deciding just how effective a given treatment is. To answer this question, one might set up several channels each successively more exacting in respect to the percent of improvement tested in the alternate hypotheses. The incoming data would then be plotted in each channel and the channel with the highest H_1 accepted would constitute a measure of the effectiveness of the treatment. As a matter of convenience, the several channels at the desired intervals could be superimposed on the same graph as in Figure 2. Sampling and plotting would continue until an H_1 and an H_0 are accepted in adjacent channels, or until a limiting P value is accepted.[2]

In the previous example, the hypothesis that treatment B results in 60 percent improvement was rejected in favor of the alternate hypothesis that treatment B results in 70 percent improvement. At this point, sampling was discontinued because the question of accepting treatment B as the treatment of choice was answered. If now a channel were to be based on $H_1 : P_1 = .80$ and $H_0 : P_0 = .70$, and sampling continued from the point at which the first decision was reached, it is possible that the 70 percent hypothesis (H_0) would be rejected in favor of the 80 percent hypothesis (H_1). But then, if we continued plotting within a channel testing $H_1 : P_1 = .90$ against $H_0 : P_0 = .80$, we might find that $H_1 : P_1 = .90$ is rejected in favor of $H_0 : P_0 = .80$ (see Figure 2). If assessment intervals employed above were adequate for our purposes, we would at this point discontinue further testing having accepted the hypothesis of 80% improvement in the channel where $H_1 : P_1 = .80$ and $H_0 : P_0 = .70$, and having rejected the hypothesis of 90% improvement in the channel where $H_1 : P_1 = .90$ and $H_0 : P_0 = .80$.

[2]Operating characteristics have not yet been derived for decisions based upon such multiple channel analyses. Thus in the applications described in the second and third examples, exact α and β values cannot be attached to the final decision. These procedures nevertheless provide useful estimates.

3. A test index has been devised which is purported to discriminate schizophrenic from non-schizophrenic patients. In order to determine how well the index will discriminate in a given clinical setting, a set of superimposed sequential analysis channels can be used as in example 2. Sampling would be undertaken according to a scheme that would reflect normal base rates for the various diagnostic groups. The direction of each channel entry would depend upon the correspondence of the test index classification (schizophrenic, non-schizophrenic in this case) with the criterion classification for each sampled patient. Each correct match would be entered as a unit advance in the vertical direction, while each non-corresponding pair would be entered as a unit advance in the horizontal direction. The power of discrimination of the test index would be reflected in the channel with the highest H_1 accepted.

4. An intelligence test has been introduced which can be administered and scored in a fraction of the time required for standard tests. As one approach in assessing the new test, we wish to find out how well it will predict achievement on the Wechsler-Bellevue. In this case, two aspects of the power of prediction might be considered: (1) the proportion of cases correctly predicted, and (2) the range of tolerance within which a given prediction will be considered correct. Let us say that we demand of the new intelligence test a correct prediction in more than 80 percent of the individuals tested. The criterion of how well the new test will predict Wechsler-Bellevue achievement would then be the minimum range of tolerance which will result in greater than 80 per cent correct prediction. Achievement on the new test might be plotted against achievement on the Wechsler, and several tolerance bands drawn about the principal diagonal, *e.g.*, $\pm$ two IQ points, $\pm$ four IQ points, etc. Thus the proportion of correct predictions will decrease as the tolerance decreases. Correct and incorrect predictions for each tolerance band would be identified and entered separately in a sequential channel where $H_0 = .80$. The minimum tolerance which yields an "accept H_1" decision would constitute a measure of how well the new test predicts achievement on the standard test.

Comment

The following advantages of the sequential analysis method should be noted.

1. When sequential analysis is used in testing percentages, it deals directly with the classification of individual cases. This is of particular importance in the clinical situation where the percentage of mistakes in the disposition of individuals is of prime concern.

2. A sequential channel once computed may be used for any number of problems requiring decisions within the same statistical limits. Thus as various types of analyses are made, a growing library of readily accessible pre-computed channels will be built for use with future problems.

3. Once sequential channels have been computed, the task of classifying individuals according to psychological test results, diagnosis, response to treatment, etc., as well as making channel entries, can usually be performed by clerical personnel.

4. With sequential analysis, statistical testing of research data is an ongoing process which can keep pace with the influx of pertinent cases and which signals the stop of sampling at precisely the number of cases required to yield a decision in respect to the hypotheses tested. While the use of this method with incoming data has been emphasized, it is, of course, also possible to test hypotheses by sampling from existing files.

5. For the hypotheses and error risks stated, sequential analysis will yield a decision with fewer cases on the average than other techniques will require.

References

1. Columbia University. Statistical Research Group. *Sequential analysis of statistical data: applications.* (Rev. Ed.). New York: Columbia Univer. Press, 1945.
2. Dixon, W. J., and Massey, F. J. Jr., *An Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
3. Fiske, D. W., and Jones, L. V. Sequential analysis in psychological research. *Psychol. Bull.*, 1954, 3, 264-275.
4. Wald, A. *Sequential analysis.* New York: Wiley, 1947.

AN APPROXIMATE TABLE FOR DETERMINING SAMPLE SIZE BY MEANS OF THE SIGN TEST

MOSS L. RAWN

Creedmoor Institute for Psychobiologic Studies
Queens Village, N. Y.

When the form of a distribution is not known or suspected to be other than normal, a conservative and approximate test for significance may be made with the sign test for which the only underlying assumptions "are that the differences are continuously distributed and independent" (1, p. 431). It may also be used for normal distributions although more precise methods are preferable. For those experimental

TABLE 1. TWO TAIL SIGN TEST AT P OF .05 FOR SAMPLE SIZE AND NUMBER OF POSITIVE AGREEMENTS[1]

N	*m*	*p*%	*N*	*m*	*p*%
6	6	98.33	41	28	66.51
7	7	94.28	42	28	66.30
8	8	91.00	43	29	66.11
9	8	88.22	44	29	65.90
10	9	85.95	45	30	65.73
11	10	84.18	46	31	65.54
12	10	82.41	47	31	65.38
13	11	81.07	48	32	65.18
14*	12	79.78	49	32	65.02
15	12	78.66	50	33	64.86
16	13	77.62	51	33	64.70
17*	14	76.70	55	36	64.12
18	14	75.88	60	39	63.50
19	15	75.15	65	41	62.92
20	15	74.40	70	44	62.44
21	16	73.76	85	53	61.22
22	17	73.18	100	61	60.30
23	17	72.65	125	74	59.16
24	18	72.08	150	88	58.34
25	18	71.60	175	101	57.69
26	19	71.15	210	120	57.00
27	20	70.74	300	168	55.82
28	20	70.32	400	221	55.02
29	21	69.79	500	273	54.48
30	21	69.56	625	338	54.00
31	22	69.22	1,200	635	52.87
32	23	68.90	2,500	1,300	52.00
33	23	68.60	5,000	2,570	51.40
34	24	68.29	10,000	5,099	50.99
35	24	68.00	40,000	20,197	50.49
36	25	67.72	1,000,000	500,981	50.09
37	25	67.45			
38	26	67.21			
39	27	66.97			
40	27	66.72			

[1] N is sample size.
m, the *larger* part of N, is the number of positive agreements rounded off to the next highest integer.
p% is equal to *un*rounded $m/N \times 100$. Minor discrepancies would exist, because of rounding, between table *p*% values and less accurate *p*% values that would derive from the listed *m* rounded values.

* A minor correction may be subtracted, although this is hardly necessary. See text for comment.

purposes where a quick and approximate way of determining sample size is needed for a given number of positive agreements, the sign test may be employed. Walker and Lev, regrettably do not make available a table for ready, easy reference which may be constructed using their sign test formula [1, p. 431]. Solving for m, *larger* part of N sample, number of positive agreements, with P at .05, yields:

$$m = \frac{1.96\sqrt{N} + N + 1}{2}$$

No calculations from the formula are needed to use table 1 except for interpolated values.

Frequently the question arises about size of sample needed for significance with a given proportion of cases. One may wish to know, for example, how large a sample of shoppers is needed for significance at .05 with 65% of the sample expected to be women. In the p% column 64.86 is found* to correspond to an m of 33 and an N of 50. Therefore 50 cases would be required to anticipate 65% of them to be women.

Another type of question is as follows: if 150 patients received medication, how many should show improvement to defeat the null hypothesis at .05? The table shows that 88 persons would have to improve for significance to be demonstrated (for more precise purposes, the corresponding p% figure 58.34 may be used).

Use of the sign test for less than 25 cases instead of the generally recommended [1] cumulative binomial probability distribution, results in very little loss of precision. This loss is in a conservative direction, *viz.*, sign test calls for *larger* m values to accord significance than does the binomial approach. Accordingly, in the table for the items marked with an asterisk N of 14 and 17, the corresponding sign test table rounded m values 12 and 14 should be reduced to the binomial values 11 and 13 respectively. All other table m rounded values need no correction. The *un*rounded p% figures found in the table, however, do not correspond exactly to binomial values. For rigorous purposes, the latter approach should be used for less than 25 cases, although under ordinary circumstances the sign test should serve these small samples well enough.

Reference

1. Walker, H. and Lev, J. *Statistical Inference.* New York: Henry Holt, 1953.

*Rounding always to next *highest* integer or 65.

CALL FOR PAPERS

This Journal is now able to offer immediate publication for high priority research reports. Increased efficiency of publication processes now permits publication of acceptable papers in three months from date of receipt. Submit your studies now.

CONSTRUCTION OF A TABLE OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN VERBAL AND PERFORMANCE IQ'S ON THE WAIS AND THE WECHSLER-BELLEVUE

WOLF P. WOLFENSBERGER[1]

George Peabody College for Teachers

Problem and Method

Clinicians often assign significance to a wide discrepancy between Verbal and Performance IQ on the Wechsler Scales. However, there appears to be lack of agreement as to the discrepancy necessary to warrant special consideration. It is felt that a table eliminating cumbersome computations in this area and that affords a rapid and objective inspection of significance values would be very helpful. The technique of Payne and Jones[(1)] was used to obtain a standard score Z, one-tailed and two-tailed probabilities for the differences between a Verbal and a Performance IQ.

Table 1 gives these data for both the WAIS and the Wechsler-Bellevue.

Table 1. Significance of the Difference Between Verbal IQ and Performance IQ on the WAIS and the Wechsler-Bellevue

Difference (d) between the Verbal and Permance IQ	Z Score		Probability ofobtaining or exceeding d		Probability of either Verbal or Performance IQ to exceed the other by this d	
	WAIS	W-B	WAIS	W-B	WAIS	W-B
8	0.82	0.70	.206	.242	.103	.121
10	1.03	0.88	.152	.189	.076	.095
12	1.23	1.05	.109	.147	.055	.074
14	1.44	1.23	.075	.109	.037	.055
16	1.64	1.40	.050	.081	.025	.040
18	1.85	1.58	.032	.057	.016	.029
20	2.05	1.75	.020	.040	.010	.020
22	2.26	1.93	.012	.027	.006	.013
24	2.46	2.11	.007	.017	.003	.009
26	2.67	2.28	.004	.011	.002	.006
28	2.87	2.46	.002	.007	.001	.003
30	3.08	2.63	.001	.004	.000	.002
32	3.28	2.81	.000	.002	.000	.001
34	3.49	2.98	.000	.001	.000	.001

Clinicians and research workers are cautioned to accept the values offered in the table as being of strictly statistical derivation and significance. The values for the Wechsler-Bellevue are less reliable than for the WAIS, because of the variations in means and standard deviations at different age levels.

Summary

In order to rapidly and objectively determine the significance of the difference between an individual's Verbal and Performance IQ on the WAIS and the Wechsler-Bellevue, a table was constructed listing Z scores and probability levels for a range of differences.

References

1. Payne, R. W. and Jones, H. Gwynne. Statistics for the investigation of individual cases. *J. clin. Psychol.*, 1957, 13, 115-121.

[1]This study was completed when the author was a Clinical Psychology Trainee at the Norfolk (Nebraska) State Hospital.

THE EFFECT OF A RECORDED VOICE ON THE DRAWING OF A GEOMETRIC DESIGN AS A PREDICTOR OF SOCIAL BEHAVIOR*

HOWARD J. SHEAR

Wilmington VA Hospital

PROBLEM

This is an exploratory study of a systematic situational testing technique to predict social behavior. The technique is based on Starer's [1] study which concluded that affective stimuli in the form of personal questions measurably affected subjects' drawings of Bender Gestalt designs. Starer did not isolate a particular affect, nor did he control adequately for nonexperimental differences in the subjects' capacities to improve in their drawing of the designs, but it did seem to lend some degree of support to the hypothesis that personally threatening affect measurably affected subjects' drawings of geometric designs. In order to further clarify this hypothesis and evaluate the use of a modified version of this technique to predict social behavior, the following hypotheses were tested: First, N-P patients will improve less in their drawing of a geometric design under the stimulus of a recorded voice of a relevant authority figure than will hospital aides, improvement being measured by decrease in time and improvement in quality from first to second performance. Second, rated cooperativeness of the subjects will be related to the amount of improvement in their drawing of the geometric designs under the stimulus of a recorded voice of a relevant authority.

METHOD

Two groups of fifteen male subjects each were studied. The groups differed on the criterion of behaviorally being presently able or unable to adjust to social living. Holding a job as a hospital aide in a VA General Medical or Surgical Hospital or being a psychiatric patient in the same hospital were considered prima facie evidence of falling in one group or the other. All of the aides in the hospital were ranked by the chief nurse on their cooperativeness with authority. The patients were rated on their cooperativeness with authority by 6 aides making independent ratings. Measures of concordance and correlation revealed a surprising amount of agreement between aides in their rankings. Separating the aides into two groups revealed a rho for the rankings of .83. Five persons from upper, middle, and lower parts of the rankings were selected as subjects.

Each subject was tested individually. They were presented 12 geometric designs, one at a time. Structurally, these designs were one basic design, composed of 4 component forms, rotated in twelve different positions in order to have 12 presentations. Immediately after drawing the designs, the subjects were given the following instructions: "That was very good. I would like you to draw these twelve designs again, but this time as you draw them you will be listening to a record. As you listen to the record, please draw the designs. Do you have any questions?" As the subject drew the design for the second time, a record was played on which the assistant manager of the hospital made a statement regarding hospital discipline, a clerk in the hospital read a column from Life magazine about European nobility, and the arm of the record player was lifted from the record. There was a balanced presentation of these three stimulus conditions, so that each subject drew 4 designs under each condition.

Time to complete each design was measured, and each design was scored as to quality. Each of the four component forms, the relationship between them, and the design as a whole were judged on a five point scale of very good to very bad. These ratings were summed to obtain a quality score for each design. A three week test-retest scoring reliability study revealed a Pearson product moment r of .93 for quality scores. Differences between first and second performances were calculated, and an analysis of variance design was used to test hypothesis 1 and correlational techniques were used to test the second hypothesis.

*I would like to thank M. W. Gasper, M.D., Manager, Miss O. L. Hufcut, Chief Nurse, and members of the N-P and Special Services of the Wilmington VA Hospital for their cooperation in this study.

Results

The data in Table 1 indicate that the patient group was more heterogeneous with regard to the quality of their initial drawings and the time needed to complete these drawings than was the aide group. Apparently, the present technique can measurably discriminate between the two groups.

Table 1. Differences Between Patient and Aide Groups with Regard to the Mean Quality and Time Scores on the First Performance

Factors		Mean	Variance	t	F
Quality	Aides	28.71	10.96	.96	2.48*
	Patients	27.13	26.62		
Time	Aides	33.4	137.62	1.13	7.41**
	Patients	43.37	1020.53		

*Significant at .05 level
**Significant at .01 level

Discussion

The data in Table 2 seem to indicate that hypothesis 1 may be rejected. There was no apparent difference between N-P and aide groups with regard to their improvement in quality and time of drawing the designs. There were no differences

Table 2. F Ratios for Patient and Aide Groups, Voice of Authority, Neutral Voice, and No Voice Conditions, and Interaction, Based Upon Quality and Time Scores

	Quality	Time	Required F at .05 Level
Subjects	.31	.71	2.53
Voices	.08	.35	19.48
Interaction	1.31	1.12	3.11

between voices of authority, neutral voice, and no voice stimulus conditions in this regard, nor were there apparent interaction affects between these stimulus conditions and the two groups of subjects. Correlations were obtained of .72 and .47 between patients' ranking in cooperativeness and reduction in time needed to complete the designs under the stimulus of the voice of authority and the more neutral voice respectively. Apparently, the more cooperative the patient was judged the less he decreased in time needed to draw design from first to second performance. This does not seem to indicate a difference in the two stimulus conditions however, for there was no statistical difference between these two correlations, t for corresponding Zs is .97.

The data seem to indicate that hypothesis 2 may be partly accepted. Although the correlations between judged cooperativeness and improvement in quality all seemed to be insignificant, the correlation of .72 discussed above suggests that the greater the judged cooperativeness of the patients the less they decreased in time necessary to draw the designs under the stimulus of the voice of the assistant manager from first to second performance. An r of $-.38$ for this same relationship with aides, although not statistically significant, suggests an opposite tendency with the aide group. It should be noted here that judged cooperativeness with the aides was determined by the ranking of the supervisory nurse who had less day to day contact with the aides than did the 6 aides who ranked the patients.

The present study seems to reveal a lack of potency of the independent variable, a record of the voice of a particular assistant manager, as well as the probability that the two groups differing in their distribution of time and quality scores in initially drawing the designs, had comparable capacities to improve in their drawing

of the designs on the second trial. The present finding of equivalence between the three stimulus conditions plus the significant relationship found between patients judged cooperative and decrease in time between first and second trials needed to draw the designs, seem to suggest that a similar analysis of variance design, using a more potent independent variable might differentiate important relationships which might aid in the prediction of social behavior.

Summary

A group of N-P patients and a group of hospital aides drew 12 geometric designs. Immediately after their initial drawings, they drew the designs again under a balanced presentation of three stimulus conditions. A recording of the assistant manager of the hospital making a statement regarding hospital discipline, a recording of a clerk reading a passage from Life magazine, and no voice presentation were the three stimulus conditions. Amount of change in quality and time from first to second performance were the dependent variables.

An analysis of variance design revealed that there were no differences between the N-P and aide groups with respect to their improvement in their drawing of geometric designs. A product moment correlation of .72 suggested that the greater the judged cooperativeness of the patients, the less they decreased in time necessary to draw the designs under the stimulus of the voice of the assistant manager from first to second performance. The r of $-.38$ for this same relationship with aides, although not statistically significant, suggests an opposite tendency with the aide group.

References

1. Starer, E. The Effects of two simultaneous cognitive and affective stimuli. *J. clin. Psychol.*, 1952, 8, 402-405.

A NOTE ON THE RELIABILITY OF THE L-M FERGUS FALLS BEHAVIOR RATING SCALE

LLOYD K. SINES

Fergus Falls State Hospital, Minnesota

Introduction

The L-M Fergus Falls Behavior Rating Scale [3, 4] has been used as a criterion measure in a number of published studies dealing with the effects of specific therapies on various psychiatric samples. [7, 5, 2] It is obvious that the efficacy of this procedure depends upon the reliability and validity of the rating scale itself. Although both of these aspects of the scale were dealt with in the earlier reports [3, 4], the present paper deals with certain limitations in the use of the scale because of the characteristics of subscale reliability.

Procedure

Sixty seven psychiatric patients of varying diagnoses (preponderantly schizophrenia) who were currently in some phase of a *total push* program [6] at the Fergus Falls State Hospital were rated on the L-M scale by four psychiatric aides who were well acquainted with them. The L-M ratings were done in the usual manner, none of the raters communicating with any of the others in regard to specific patients or ratings. Of the total group of patients, 33 were male and 34 were female. Two male aides rated the male patients, while two female aides rated the female patients. Area F on the scale (Response to Electric or Insulin Therapy) was not rated since none of the patients had received either therapy during the period on which the ratings were based.

Results

Several measures of reliability were computed. First, since for each patient two L-M ratings were obtained, an *inter-rater* reliability coefficient was obtained. That is, aides A and B rated (male) patients 1 through 33 from which r_{AB} was obtained by correlating the two mean L-M scores for X_1 (patient number one) through X_{33} (patient number 33). Similarly, aides C and D rated (female) patients 34 through 67 from which r_{CD} was obtained by correlating mean L-M scores for X_{34} through X_{67}. Finally, the male and female groups were combined and an overall reliability coefficient ($r_{I\ II}$) was computed between the 67 L-M means obtained from the four raters (Table 1).

Table 1. Parameter Values and Inter-rater Reliability Coefficients Obtained by Correlating Mean L-M Scores on a Sample of 67 Intensive Treatment Patients

Sample	Average L-M Scores			Inter-rater reliability
	N	Mean	S.D.	
Male	33	2.97	.608	.843
Female	34	2.89	.800	.894
Total group	67	2.93	.714	.853

As a further estimate of inter-rater reliability, the per cent of agreement between raters on subscale ratings was computed (Table 2). It may be seen from Table 2 that in 92 per cent of the ratings the two raters either were in absolute agreement or differed in their ratings by only one point.

Table 2. The Per Cent of Agreement Between Raters and the Number of Cases in Which Discrepancies of Varying Amounts Obtained Between the Two Raters

Discrepancy between raters	N	%
0	373	56.7
1	233	35.4
2	43	6.5
3	9	1.4
4	0	0.0
	658	100.

The values reported in Tables 1 and 2 compare favorably with those reported in the original article on the scale[3] and indicate that the inter-rater agreement, in terms of *total* L-M score and per cent of agreement on actual ratings, is adequate to the demands of most clinical-type research in which the scale is used. However, demonstrating the reliability of *total* or *mean* scores does little in the way of establishing the degree of reliability of the various *subscales* or *areas* included in the scale as a whole. Since it has been suggested that pre-post therapy analyses of subscale scores may elucidate specific areas of change in patient behavior, a determination of the reliability of the individual scales is in order. The following procedures were carried out in obtaining the pertinent values. Each of the ten L-M areas (subscales) rated was considered individually and inter-rater reliability coefficients computed. Thus r_{AB}, r_{CD}, and $r_{I\ II}$ were obtained for each of areas A through K, with the exception of area F (Table 3). From Table 3 it is apparent that the ten subscales differ with respect to reliability, and with possibly one or two exceptions, none approaches the level of reliability usually demanded of complete tests.

TABLE 3. THE INTER-RATER RELIABILITY COEFFICIENTS OF TEN OF THE L-M SUBSCALES (AREAS)

Area	Male sample				Female sample				Total sample			
	N	Mean	S.D.	r_{AB}	N	Mean	S.D.	r_{CD}	N	Mean	S.D.	$r_{I\ II}$
A	33	2.64	.648	.773	34	2.91	.812	.779	67	2.73	.825	.772
B	33	3.32	.812	.577	34	3.21	.880	.707	67	3.22	.922	.656
C	33	2.68	.860	.652	34	2.14	1.086	.810	67	2.37	1.054	.759
D	33	3.06	.872	.637	34	2.86	1.126	.675	67	2.92	1.058	.667
E	32	2.75	.566	.563	33	2.85	.764	.697	65	2.80	.671	.616
G	33	2.47	.883	.396	33	2.55	1.013	.599	66	2.51	.975	.524
H	32	3.02	.949	.804	33	2.61	1.013	.689	65	2.81	1.000	.655
I	32	3.36	.872	.730	33	3.15	1.281	.410	65	3.25	1.018	.513
J	31	2.97	1.100	.622	33	3.17	1.039	.786	64	3.07	1.182	.641
K	32	3.34	.825	.705	33	3.67	.823	.408	65	3.51	.825	.551
A through K	648	2.96	.806	.661	668	2.88	1.118	.663	1316	2.92	1.010	.653

Discussion

The apparent lack of high subscale reliability precludes the feasibility of using specific L-M items or areas as *tests* by themselves. Further, pre-post therapy analyses of specific areas on the scale are hazardous in view of the relatively low reliability of several of the subscales. For example, apparent changes in psychomotor activity (area I) ratings after therapy could arise from the low reliability of that scale as well as from an actual effect of the therapeutic procedure on the patient's psychomotor behavior.

Although the subscale reliability coefficients range from .513 to .772 and the overall (A through K) reliability of the subscales is but .653, the substantially higher reliability of total or *mean* L-M scores indicates that the scale may be used *as a whole* with assurance that observed changes in the latter are reliable within the limits generally accepted. Finally, in view of the demonstrated validity of the scale[4], it may be said that significant changes in total or *mean* L-M scores reflect actual changes in the behavior of the patients rated.

Summary

Sixty seven psychiatric patients on an intensive treatment service of a state hospital were rated by four psychiatric aides on the L-M Fergus Falls Behavior Rating Scale. Various measures of reliability were computed on both total or mean L-M scores and the actual subscale ratings. It was found that while mean L-M scores have adequate inter-rater reliability, the several subscales differ with respect to reliability and none approaches the usual standards of adequate *test* or complete scale reliability. It was concluded that while the L-M scale can be used *as a whole*, treating total or *mean* score as a dependent variable, subscales should not be used individually, nor should specific "item analyses" be made to determine the actual areas of change in patient behavior as a result of significant total score changes. It is possible only to infer changes in "gross observable behavior" from significant L-M total score changes rather than to delineate with high accuracy specific *areas* of behavior change.

References

1. Atkinson, S., Fjeld, S. P., and Freeman, J. G. An Intensive Treatment Program for State Hospital Geriatric Patients. *Geriatrics*, 1955, 10, 111-117.
2. Freeman, J. G., and Lucero, R. J. F-L Fergus Falls Lobotomy Scale. *Arch. Neurol. Psychiat.*, 1956, 76, 30-33.
3. Lucero, R. J., and Meyer, B. T. A Behavior Rating Scale Suitable for Use in Mental Hospitals. *J. clin. Psychol.*, 1951, 7, 250-254.
4. Meyer, B. T., and Lucero, R. J. A Validation Study of the L-M Fergus Falls Behavior Rating Scale. *J. clin. Psychol.*, 1953, 9, 192-195.
5. Penman, A. S., and Dredge, T. E. Effect of Reserpine and Open-Ward Privileges on Chronic Schizophrenics. *Arch. Neurol. Psychiat.*, 1956, 76, 42-49.
6. Sines, J. O., Lucero, R. J., and Kamman, G. R. A State Hospital *Total Push* Program for Regressed Schizophrenics. *J. clin. Psychol.*, 1952, 8, 189-193.
7. Sommerness, M. D., Lucero, R. J., Hamlon, J. S., and Mahowald, A. M. Chlorpromazine: A Controlled Study with Highly Disturbed Patients. *Dis. Nerv. Syst.*, 1957, 18, 1-4.

EDITORIAL OPINION

THEORY VERSUS PRACTICE IN CLINICAL PSYCHOLOGY

There is a wide discrepancy in clinical psychology between theory and practice, between what we believe and what we do. Our tests, for instance, are at variance with our general theory of the nature of maladjustment. We state convincingly that the problem resides not in individuals, in parents or in children, but in a faulty relationship between the two. But our psychological tests were devised to learn what is wrong with the individual, thus placing the locus of the problem in him. What we need is a whole new battery of tests, or new ways of interpreting the tests we have, to get at the faulty relationship rather than the faults of individuals.

The search for the trouble in an individual turns on an antiquated theory of causality. There have been many criticisms of the traditional doctrine of cause and effect, but the one that is most pertinent here is the notion of passivity. In the mechanical view, the cause was considered to be the active agent and the effect the passive consequent. But in human relationships the so-called effect or consequent is as active as the so-called antecedent or agent, so that there is a reciprocal effect on each other and both persons are affected. Despite our disavowal of the old empirical view of causality in the practice of clinical psychology, by the use of tests we assume that the parents are the agent or the cause, and the child is the effect or the result of the interaction which naturally exaggerates the role of the parent and minimizes the role of the child. The formulation of a modern theory of causality consonant with human interaction is a desideratum.

The prevalent attitude towards symptoms, as surface signs hiding the true cause or the real illness, is related to our notion of cause and effect. Illness, in clinical psychology, is an abstraction, and only the symptoms are real, for they are the maladjustment. The search for the true cause is a search for the non-existent. This erroneous kind of thinking may be due to the analogy to medicine where an illness may have definite causes and symptoms. Emotional problems are not an illness in the same sense; they are maladjustments due to maleducation.

Another factor lies in our theory of personality. We constantly state that personality is not fixed or set like plaster, but something pliable and protean. Yet our tests assume that we have a veridical personality that can be understood by making a few soundings or plumbing of its homogenous depth. While we admit that personality is changing, we explore it as if it were stable rather than in a state of flux. We treat it as if it were a unitary thing, consistent and uniform, whereas it may be that its diversity is more important than its integration, and calls for a more suitable figure of speech that represents this construct.

Another point turns on the two opposing pillars upon which our psychology is built. One of these pillars is the psychology of individual differences: everyone is fundamentally unique and different. This is a grand generalization, but it is also a generalization to end all further generalizations: it doesn't make it possible to formulate any laws or principles and limits clinical psychology to an art rather than a science. On the other hand, we also believe that everyone is fundamentally the same. Our differences are so slight that they can all be placed on the same continuum and form a normal frequency curve. This permits the statistical treatment of personality.

The acceptance of these two diametrically opposed concepts without any attempt to harmonize them widens the gap between theory and practice, for it enables us to profess one position and practice the other. We are able to justify the use of either frame of reference for our theory or practice. Clinical psychology will continue to suffer from a disparity between theory and practice until these two basal dicta are resolved. This task seems more important to the success of clinical psychology than the verification of current theories or the formulation of new ones.

SALVATORE RUSSO
Wichita (Kansas) Guidance Center

"BROAD SPECTRUM" CLINICAL PSYCHOLOGY

As applied to clinical practice, "broad spectrum" refers to methods which encompass the entire range of factors related to etiology, diagnosis and therapy. In contrast, "narrow spectrum" methods are directed toward only some limited range or aspect of clinical phenomena. The concept recently has attained wide recognition in clinical medicine, particularly with reference to wide-range antibiotics and other drugs which are effective against a wide range of etiologic agents in contrast with earlier methods having only limited effectiveness. It is intended to emphasize the necessity of recognizing and dealing with a wide range of etiologic factors comprehensively.

There is evidence that not all viewpoints in clinical psychology are properly described as "broad spectrum" approaches. At the 1957 APA convention at New York, there was considerable joking about colleagues who persist in playing one-stringed violins. Thus we have the EENT (eye-ear-nose-and-throat) psychologists who are still preoccupied with brass instrument studies of sensation. And the "rat" psychologists or reflexologists who preach that all can be known from the behavioristic or neobehavioristic study of the laboratory animal—they paradoxically ignore the complexities of higher level phenomena such as interpersonal relations. And the dyed-in-the-wool experimentalists who sometimes sound as sad when they branch off into clinical speculations as do intuitive clinicians at play in the experimental laboratory—in fact we were positively embarrassed at the APA to hear what some noted experimentalists had to say about clinical matters when they wandered afield from their stated topics. And so on through adherents of all the other schools, each apparently deriving some inner gratifications from identifying with or creating some special theoretical position.

It is difficult to see how a genuinely "broad spectrum" approach can be anything but thoroughly eclectic. "Broad spectrum" means more than the attempt to broaden out any one theoretical model until it can explain everything in the field. Is clinical psychology a separate specialty or profession, or is it simply one application of basic science psychology? Some experimentalists appear to be in the position of arguing that anyone with a solid background in basic science psychology, *e.g.* a solid experimental-statistical orientation, should be competent to practice in the applied fields which are regarded as simply practical applications of basic science. The argument seems to be that a good experimentalist has the basic tools to succeed in any specialty he attempts. Without denying the value of a sound experimental-statistical background, we very seriously doubt whether anything can replace years of experience and clinical practice with actual case materials in our field.

If anything, the clinician has to be more "broad spectrum" eclectically-oriented than the experimentalist, statistician, comparative psychologist, psychophysiologist or learning expert. Anyone who proposes to deal with the personality as a whole, studying the organism-meeting-its-environment, and dealing with the complexity of interaction taking place in psychosocial fields, must indeed be a master scientist, familiar with and competent in all the borderline specialties which may contribute to the problem. At a time when a genuinely "broad spectrum" approach to clinical practice is possible, anyone who voluntarily limits himself to anything less broad must be regarded as not practicing according to the highest standards of time and place. The older generation of clinical psychologists may be partially excused from living up to "broad spectrum" qualifications of competence because of training deficiencies of their era. But the younger generation has an obligation to live up to the "broad spectrum" developments which are just over the horizon. The "broad spectrum" approach is essentially an attitude, a characteristic manner of thinking and operating in regard to data.

F. C. T.

BOOK NOTICES

JONES, ERNEST. *The Life and Work of Sigmund Freud.* Vols. I, II, III. New York: Basic Books, 1953, 1955, 1957. Vol. I, $6.75. Vol. II, $6.75. Vol. III, $7.50.

Perhaps never before have the fortuitous circumstances occurred making possible an intensive biographic study of an authentic genius made by a distinguished psychoanalyst utilizing a stupendous amount of data gathered from life record, introspective reporting, objective examinations and projective sources. Freud, the founder of psychoanalysis, lived in a sort of psychoanalytic fish bowl in the sense that his students and followers repeatedly scrutinized their master in a succession of penetrating analyses, most of which have gotten into print. Probably it can be said that never before has any one person been so intensively studied by so many psychoanalytic experts as was Freud. While other geniuses have been venerated and written about, it is safe to say that never before has so much intimate evidence been collected concerning the life and inner motivations of any one man. The result of all this intensive study is one of the best biographic studies ever written. This review must start by paying homage to Ernest Jones who was not only one of Freud's most staunch friends and followers, but who also must be granted recognition as a sterling scientist and writer in his own right. Having known Freud for thirty years, and labored with him unceasingly to advance the cause of psychoanalysis throughout the period, and finally being instrumental in securing refuge for Freud in England when he was endangered by the Nazi outrage, Jones may lay claim to having known the master as well as anyone not a member of the family ever could. Science will be eternally grateful to Jones for having undertaken the massive research and literary project necessary to collate so much material.

No review could be capable of summarizing the numberless insights into Freud the man which are contained in these three volumes. These the reader can only get for himself by patient reading, and rereading, the fascinating story of how Freud slowly matured into the gigantic figure which he was. Here are only a few points which intrigued this reviewer. It appears that Freud as a small lad had some intuition concerning his own genius and great destiny for there are repeated examples when he seemed inspired from an early age by inner creative spasms from which original thoughts evolved. It appears that beginning with his mother, the quality and depth of his creative genius quickly made itself apparent and won many friends and supporters. Freud early showed an intensity of feeling and thought which repeatedly manifested itself in many overt behaviors. The complexity and sensitivity of his personality were particularly revealed in the troubled years of courtship when he was still too poor and unsuccessful to get married. That these did not unsettle or distract him is a tribute to his own stability and the inner strength resulting from his personal analysis of himself. Freud was a prolific writer, and was written about prolifically, so we have a rich abundance of data concerning what he seemed to be externally and when went on introspectively. This reviewer constantly experienced feelings of a deep understanding and admiration of Freud as a man on reading the intimate details of his life. In his own personal life, Freud came as close to being as healthy, mature, stable, responsible, dedicated, understanding, tolerant and creative as any mortal could possibly be. From the emotional turmoils of courtship, through the bitter travail of his creative scientific life, to the final heroic struggle with fatal illness, the personal life of Freud gives as convincing a picture of inner strength and self-consistency and decency as can be imagined. Freud had his conflicts, ambivalences and impulsive moments, but these are tremendously overshadowed by the dedication to long range goals and sincere determination to do what was right in both personal and scientific life. Jones has patiently documented in great detail what actually went on during the critical periods of Freud's life and it all makes fascinating reading.

The story of how Freud patiently constructed the theory and practice of psychoanalysis raises him to the rank of being one of the greatest empiricists in the history of scientific discovery. Although Freud is known for his great deductive contributions, his basic reliance was on the most painstaking and thorough study of raw behavior by inductive methods. His early training in neuropathology gave him a solid grounding in the experimental-objective approach to raw data, and he went far beyond just applying the popular theories of his era to the creation of entirely new ideas. One of his greatest abilities was a fine flair for recognizing the crucial importance of critical incidents which gave clues which lesser men ignored. Repeatedly, we find him recognizing the significance of relationships which others had just passed over. He kept asking himself: "What could this mean?—and then coming up with original insights in a truly astounding profusion. The historical developments of the great discoveries of psychoanalysis are too well known to be repeated here. The truly unusual thing is the fact of such great productiveness in just one man.

It has been said that no radical thinker is ever able to hold out in his reformist efforts into old age in the face of the cumulative weight of opposition and derogatory public opinion. The penalty of original thinking usually is to be ignored and ostracized by the entrenched political forces and dogmas of the day. Freud suffered more than most productive geniuses in having his discoveries ridiculed and derided by the mass of more orthodox thinkers. His theories concerning the sexual and emotional etiology of neurotic and normal behaviors outraged the conventional morality of his times. Indeed his treatment at the hands of the Viennese authorities is but another proof of the old axiom that no man can be an hero in his own town. Lesser men would have desisted and shriveled in the face of such powerful opposition but Freud persisted steadfastly in his duties as he saw them. One of the marks of a great man is to be unceasingly active, whether right or wrong; lesser men vegetate in the safety of passivity and conventionality. One of his rewards was to live to see the day when his great discoveries would be publicly recognized and heralded for their true worth.

The critical reader will be interested to note that Freud himself remained hardheadedly critical of himself and the developments of his times to the end of his life. Never was there any weakness or deterioration, even in old age, in his determination to discover the truth and to validate his hypotheses and findings by every possible means. Freud himself was indubitably more critical than many of his followers have been, and he probably would unmistakably dissociate himself from many of the excesses which have characterized the later development of psychoanalysis. Even though its strongest propounder and defender, Freud himself was broader than psychoanalysis in that he repeatedly applied the law of parsimony and revised his own theories when he saw that new evidence demanded it. The breadth of Freud's education was truly remarkable for his day, and his insights were far broader than those of many figures on the current scene. One may quarrel with points of psychoanalytic theory but we cannot fail to recognize the soundness and depth of Freud the man.

One of the strong points of this biography is the mass of interpretive material which Jones has added at each point in the historical development of the story. Indeed, the student can learn about psychoanalysis from reading the life of Freud and how one development led to another. This reviewer feels that *The Life and Work of Sigmund Freud* must be required reading for every psychologist and psychiatrist. Next to having been intimate with Freud as a living person, the best thing is to become thoroughly familiar with this monumental objective study of what he did and was.

F. C. T.

CATTELL, RAYMOND B. *Personality and Motivation Structure and Measurement.* Yonkers, N. Y.: World Book, 1957, pp. 948.

Dr. Cattell and his University of Illinois Laboratory of Personality Assessment and Group Behavior staff have accomplished a truly stupendous amount of work in personality study using factorial methods. The development of his theory, research

program and practical test design has progressed so rapidly as to require frequent extensive reviews of their theoretical and practical implications. The fact that Dr. Cattell is one of the most prolific and original thinkers on the current psychological scene renders it very difficult to write an authoritative review of all the concepts and data included in this compendious work which covers such a wide range of topics that not many contemporary psychologists are competent to comprehend the technical details of all that is presented. Indeed, the whole field of factor analysis is so new and with so many unsettled problems, that the ultimate validity of much that Cattell has done must await the time when there can accumulate sufficient data and cross-validation experiments so that general agreements on basic issues can be reached. Cattell's sweep is so broad that his enthusiasm frequently leads him into relatively uncharted areas where speculations reach far beyond what has been actually validated. It will require patient checking in independent laboratories to determine whether Cattell's results can be replicated and cross-validated.

This reviewer does not wish to be guilty of using the device of slyly derogating a major work by citing a number of apparent inconsistencies or controversial points and thereby casting suspicion on the value of the work as a whole. Cattell's work has not received the attention it deserves in American psychology, partly because its scope is beyond the comprehension of many of his critics and partly because Cattell's enthusiasm to pioneer new areas has sometimes led him into premature conclusions which have been unfairly cited as characterizing his whole work. We suggest in all seriousness that the AMERICAN PSYCHOLOGICAL ASSOCIATION would do well to organize symposia and study groups to bring the best minds in the profession to bear on Cattell's work in order to arrive at a better understanding of its ultimate validity. In the meantime, this advanced text is heartily recommended to those serious readers who have the patience and the background to encompass all that is presented.

CORSINI, RAYMOND J. *Methods of Group Psychotherapy.* New York: McGraw-Hill, 1957, pp. 251. $6.50

The author is research associate at the Industrial Relations Center of the University of Chicago and counselor at the Chicago Community Child Guidance Centers. This is an eclectic textbook which objectively presents the current status of this new implement in the psychotherapeutic armamentarium. Part one consists of ten chapters discussing the history, nature, mechanisms, procedures, processes and applications of group therapy. Part two is a case book presenting representative verbatim recordings of analytic, nondirective, family counseling and psychodramatic group therapy.

KARPMAN, BEN. *The Hangover.* Springfield, Ill.: C. C. Thomas, 1957, pp. 531. $9.50

Dr. Karpman, one of America's foremost authorities on problems of psychopathology, here approaches the psychological problems of alcoholism by a detailed study of the hangover. This is accomplished empirically and inductively by presentation of a detailed series of case histories illustrating most of the known phenomena of alcoholism. The orientation is psychoanalytic and the presentation so authoritative that this book will undoubtedly take its place as a basic source book on the subject.

WOLFF, MAURICE D. and WOOLF, JEANNE A. *Remedial Reading.* New York: McGraw-Hill, 1957, pp. 424. $5.75

This is one of the best texts on remedial reading which we have seen. The problem of remedial reading is approached within the broad frame of reference of personality as a whole, citing the importance of making the most detailed psychological studies of general personality development against which reading problems and other symptoms may be evaluated. Theory and practice are presented eclectically, with well organized illustrative materials.

SPRING LAKE RANCH

A Farm in the Green Mountains Where Men and Women Facing Difficult Life Problems May Build Health, Confidence and Courage.

Spring Lake Ranch offers an experience in normal group living to men and women well enough to cooperate in their own behalf. It is no sanitarium, but the joint outgrowth of the faith, leadership and skills of founders, staff and guests.

WHAT SPRING LAKE RANCH HAS TO OFFER—a free natural environment, and permissiveness that is unique and valuable, rural family life, all ages, both sexes, a self-governing group, in which guests help to plan and direct activities, a realistic program of work and play, emphasizing the big muscles, a mature plan leading to self-support as an incentive to adjustment. Formal psychotherapy, may be arranged for, with Dr. Frederick C. Thorne, who is also available as consultant on psychological and psychiatric problems.

WHAT SPRING LAKE RANCH DOES NOT OFFER—a "rest cure", custodial care for individuals unable to cooperate, resident medical or psychiatric care, full-time supervision.

A LAYMAN'S EFFORT—From the start, leading psychiatrists have identified themselves with this experiment, referring suitable patients, furnishing clinical histories, advising the staff. This teamwork has made possible a center dedicated to mental health in a remote setting without a resident psychiatrist—a layman's complement to psychiatry.

REFERRALS—Ranch life is designed only for those whose constructive forces may be reached. Space being limited, we only accept those likely to improve. Results have been good with conditions of fear, anxiety, depression, obsession; pre-hospital and middle-age emotional problems; alcoholics seeking help; epileptics whose convulsions are under control; post-shock and selected post-lobotomy patients.

DURATION OF STAY—"Ranchers" are in two categories: those who require a strong pick-up, and "pressure-off" for 1 to 3 months, those who need rehabilitation between hospital and home, 3 months to 2 years.

TRAINEES—The experience of 24 years has convinced us that this type of interim facility is needed throughout the United States. We are ready to receive as trainees qualified couples, to work with us a year or more with a view to starting such centres themselves. We offer a free, good life, a small salary, staff member status and the benefit of our total experience.

WAYNE ANTHONY SARCKA :-: :-: **ELIZABETH MAN SARCKA**

Cuttingsville, Vermont

NEW AND UNIQUE - - -

THE HANGOVER

A Critical Study in the Psychodynamics of Alcoholism.

By
BENJAMIN KARPMAN, M.D.
Chief Psychotherapist
St. Elizabeths Hospital
Washington, D. C.

"In the entire set of reactions which appear to follow the use of alcohol, none, it seems, is more remarkable than that which goes under the name 'HANGOVER.'" — *From the Preface*

A frank and penetrating study of the lives of seven men and seven women alcoholics in whom the "hangover" is shown to be **symptomatic of the specific underlying personality disorders** which lead to alcoholism. Dr. Karpman's study shows that analysis of the hangover yields valuable clues as to the nature of the personality maladjustment. Individual differences in the reaction to alcohol are clearly shown in the hangover period.

REFLECTED IN THE HANGOVER ARE VIRTUALLY ALL PSYCHOLOGICAL ASPECTS OF ALCOHOLISM

Much of the material is in the words of the men and women concerned and they discuss freely the interplay of problems connected with family, sex, employment, and social life. The material presented will be valuable to workers in many fields of the social sciences as well as to psychiatrists.

Unusual original drawings by Wesley R. Wilken vividly illustrate major points in each case and convey the sufferings and the utter futility of the life of the confirmed alcoholic.

560 pages (6 x 9) **16 illustrations**
Published fall 1957 **Sent on approval, $9.50**

PARAPSYCHOLOGY

Frontier Science of the Mind

By
J. B. RHINE, Ph.D.,
and
J. G. PRATT, Ph.D.
Psychology Laboratory of Duke University

A SURVEY OF THE FIELD, THE METHODS, AND THE RESULTS OF ESP AND PK RESEARCH

Today students in the colleges want to find out what is known about *psi capacity*, and even a few courses on parapsychology are being announced. To fill this need the authors have written a book that in itself introduces the reader to the field of parapsychology, defines the terms, describes the methods, and summarizes all the main facts accumulated to date.

Provides for the convenience of professional and research workers the basic techniques used in the tests on which most of the psi investigations have depended; and adds tables and measures for their appraisal of test results.

It is **READABLE**;
it is **COMPREHENSIVE**;
it is **USEFUL** as a manual.

Written in the world center of experimental research in the field it covers, the Parapsychology Laboratory of Duke University.

232 pages (6 x 9) **10 illustrations**
Published fall 1957 **Sent on approval, $4.75**

CHARLES C THOMAS - PUBLISHER
301-327 East Lawrence Avenue
Springfield, Illinois

DEVEREUX SERVES

A full educational program, extending from kindergarten through the junior college level, is offered at Devereux for emotionally disturbed, as well as slow-learning children.

Instruction is highly individualized and planned to take advantage of the specialized learning techniques which are most effective for these children.

The curricula of the Devereux units include a solid academic program for college preparatory work; a vocational and commercial program; and programs of artisanship or craftsmanship, stressing the use of special tools and skills, either as preparation for admission to a vocational or trade school or as a means of attaining self-confidence and personality development.

Professional inquiries should be addressed to John M. Barclay, Director of Development, Devereux Schools, Devon, Pennsylvania; western residents address Keith A. Seaton, Registrar, Devereux Schools in California, Santa Barbara, California.

THE DEVEREUX FOUNDATION

A nonprofit organization — *Founded 1912*
Santa Barbara, California — *Devon, Pennsylvania*

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX, Founder
EDWARD L. FRENCH, Ph.D., Director

Professional Associate Directors
Robert L. Brigden, Ph.D.
Charles M. Campbell, Jr., M.D.
Michael B. Dunn, Ph.D.
Fred E. Henry, S.T.D.
J. Clifford Scott, M.D.